THE ANNALS [illegible]

OF

[illegible]

STAT[illegible]

[illegible]

THE OFFICIAL [illegible]

OF MATHEMATICAL [illegible]

Contents

Vol. 26, N[illegible]

THE A[illegible]NALS OF MATHEMA[illegible] STATISTICS

[illegible] quarterly by the I[illegible] Mathematical S[illegible] in [illegible] June, September and [illegible] at Baltimore, [illegible]

INSTIT[illegible] OF MA[illegible]CAL ST[illegible]

To June 30, 1956

Treasurer's Office — Department of Mathematics, [illegible] State [illegible], East Lansing, Mich.
K. J. Arnold, Treasurer

From July 1, 1956

Department of Statistics, [illegible], Stanford, Cal.
A. H. Bowker, Treasur[illegible]
These addresses should [illegible] communications concerning subscriptions, [illegible] numbers, payment [illegible], etc.

To June 30, 195[illegible]

Secretary's Office — Department of Mathematics, Mich[illegible] State College, East Lansing, Mich.
K. J. Arnold, Secret[illegible]

From July 1, [illegible]

Department of Stat[illegible] North Carolina, Chap[illegible] N. C.
George E. Nicholson, Jr., Secret[illegible]
These addresses sho[illegible] used [illegible] concerning membership, changes of ad[illegible], etc. C[illegible] are to become effective for a given [illegible] of the [illegible] should be reported to the Secretary on or before the [illegible] month preceding the month of issue.

Editorial Office — Statistical Laborat[illegible] of California, Berkeley [illegible]
E. L. Lehmann, Ed[illegible]
Manuscripts shou[illegible] to this address; each manuscript should be typew[illegible] wide margins, and [illegible] original copy shou[illegible] with one additional [illegible]). Dittoed and mim[illegible] acceptable if they are completely legible and printe[illegible] make corrections. Footnotes should be reduced to a m[illegible] possible replaced by [illegible] at the end [illegible] in foot[illegible] should be avoided. Figures, charts, [illegible] plain white paper or tra[illegible] they are to be printed. Authors [illegible] typographical difficulties of complica[illegible] Authors will ordinarily receive only galley p[illegible] without covers will be furnished free. Additional repr[illegible] furnished [illegible]

Subscription — Current issues are $1[illegible] (four issues [illegible] calendar year) in the U. S. and Canada; $10 [illegible] elsewhere. [illegible] issues are $3.50. Back numbers are $1[illegible] volume, $3.50 per issue, [illegible] for the first 25 volumes, $10 per additional volume purchased at the same time as Volume I through XXV.

Composed [illegible] at the
WAVERLY PRESS, Inc., [illegible], Maryland, U. S. A.

[illegible] second-class matter at the Post [illegible] act of March [illegible]

ON TESTS OF NORMALITY AND OTHER TESTS OF GOODNESS OF FIT BASED ON DISTANCE METHODS

By M. Kac, J. Kiefer,[1] and J. Wolfowitz[2]

Cornell University

Summary. The authors study the problem of testing whether the distribution function (d.f.) of the observed independent chance variables $x_1, \cdots, x_n$ is a member of a given class. A classical problem is concerned with the case where this class is the class of all normal d.f.'s. For any two d.f.'s $F(y)$ and $G(y)$, let $\delta(F, G) = \sup_y |F(y) - G(y)|$. Let $N(y \mid \mu, \sigma^2)$ be the normal d.f. with mean μ and variance σ^2. Let $G_n^*(y)$ be the empiric d.f. of $x_1, \cdots, x_n$. The authors consider, inter alia, tests of normality based on $v_n = \delta(G_n^*(y), \;\; N(y \mid \bar{x}, s^2))$ and on $w_n = \int (G_n^*(y) - N(y \mid \bar{x}, s^2))^2 \, d_y N(y \mid \bar{x}, s^2)$. It is shown that the asymptotic power of these tests is considerably greater than that of the optimum χ^2 test. The covariance function of a certain Gaussian process $Z(t)$, $0 \leqq t \leqq 1$, is found. It is shown that the sample functions of $Z(t)$ are continuous with probability one, and that

$$\lim_{n \to \infty} P\{nw_n < a\} = P\{W < a\}, \quad \text{where} \quad W = \int_0^1 [Z(t)]^2 \, dt.$$

Tables of the distribution of W and of the limiting distribution of $\sqrt{n}\, v_n$ are given. The role of various metrics is discussed.

1. Introduction. Let $x_1, \cdots, x_n$ be n independent chance variables with the same cumulative distribution function $G(x)$ (i.e., $G(x) = P\{x_1 < x\}$) which is unknown to the statistician. It is desired to test the hypothesis that $G(x)$ is a normal distribution. This is an old problem of considerable interest which has received a fair share of attention in the literature.

A commonly used test consists essentially in testing whether the third moment of $G(x)$ about its mean is zero and whether the ratio of the fourth moment about the mean to the square of the second moment about the mean is three. It is obvious that this is not really a test of normality, because there are many non-normal distributions which satisfy these conditions on the moments.

Perhaps the best of the commonly used large sample tests of normality is the χ^2 test due to Karl Pearson; see for example Cramér [1], Sections 30.1 and 30.3, and the recent results of Chernoff and Lehmann [2]. The asymptotic power of the χ^2 test was studied by Mann and Wald [3]. (It is true that these authors

Received August 30, 1954, revised January 10, 1955.

[1] Research under contract with the Office of Naval Research. The computations of the tables of the present paper were performed under the auspices of the Office of Naval Research to whom the authors are greatly obliged.

[2] The research of this author was supported in part by the United States Air Force under Contract No. AF18(600)-685, monitored by the Office of Scientific Research.

studied the problem of goodness of fit for a simple hypothesis, which in our problem would correspond to knowing the mean and variance of $G(x)$. However, it is plausible that the comparison in power which we will make below will be true a fortiori for our problem.) One of their principal results is the following. Define the distance $\delta(H_1, H_2)$ between any two distribution functions H_1 and H_2 by

$$\delta(H_1, H_2) = \sup_x |H_1(x) - H_2(x)|.$$

Suppose one tests the null hypothesis that $G(x) \equiv R(x)$, where $R(x)$ is a given continuous distribution, at some fixed level of significance. In [3] it is shown that the χ^2 test based on n observations and k_n intervals of equal probability under R, where k_n is chosen for each n so as to minimize the value Δ_n for which the minimum of the power function among alternatives R^* with $\delta(R^*, R) \geqq \Delta_n$ is $\frac{1}{2}$, gives $\Delta_n \doteq cn^{-2/5}$ when n is large. In Section 5 we shall show that the result of Mann and Wald just stated also holds if δ is replaced by the measure of discrepancy γ introduced in Section 5.

Let

$$\psi_x(a) = \begin{cases} 0, & x \leqq a; \\ 1, & x > a. \end{cases}$$

Let

$$G_n^*(x) = \frac{1}{n} \sum_{i=1}^{n} \psi_x(x_i)$$

be the empiric d.f. of $x_1, \cdots, x_n$, that is, $G_n^*(x)$ is the proportion of x_i's less than x. The asymptotic distribution of $\delta(G, G_n^*)$ was found by Kolmogoroff [4] and it is known that $\delta(G, G_n^*)$ is of the order $n^{-1/2}$ in probability (that is, with probability arbitrarily close to one). If now one bases the test of the hypothesis that $G(x) \equiv R(x)$ on $\delta(R, G_n^*)$, with large values of δ significant, it follows that, for large n, it is sufficient that $\delta(R, R^*)$ be of the order $n^{-1/2}$ for the probability of rejecting the null hypothesis to be appreciable ($\geqq \frac{1}{2}$). To put matters a little differently: Let $\delta(R, R^*) = h$ be small (so that n has to be large in order to distinguish between R and R^*). Then, if n has to be equal to N in order to guarantee that the power at R^* of the χ^2 test of goodness of fit be at least $\frac{1}{2}$, when one uses the test based on δ it is enough that n be of the order of $N^{4/5}$. This is a considerable improvement (for large n). The result just stated holds also for the classical "ω^2" test if δ is replaced in the above by γ; this follows from the results of Section 5.

Let us return to the problem of testing the composite null hypothesis whether $G(x)$ is normal (its mean and variance being unknown). One of the present authors has been developing the minimum distance method for estimation and testing hypotheses in a number of papers (Wolfowitz [5], [6], [7], [8]). In accord with this method it is proposed in [5] (page 149) that this test of normality be based on $\delta(G_n^*, N^{**})$ with the large values critical. Here N^{**} is the class of

all normal distributions, and the distance of any d.f. $H(x)$ from the class N^{**} is given by

$$\delta(H, N^{**}) = \inf_{N \epsilon N^{**}} \delta(H, N).$$

In Sections 2, 3, and 4 we investigate tests based on various "distance" criteria, constructed in the spirit of the above discussion, and in Section 5 we discuss the asymptotic power of these tests. As stated in [5], the minimum distance method is not limited merely to testing normality, and in Section 4 we discuss its application to other tests of goodness of fit.

Throughout this paper we discuss tests, computations, and power considerations in terms of particular (normal or rectangular) examples, but it will be obvious that the results of Sections 2, 3, and 5 may in general be carried over to testing composite hypotheses involving parametric families. The minimum distance method applies in principle to even more complicated families of distribution functions about which one desires to test a hypothesis. For hypotheses of a more complicated nature than our examples, there will often exist the additional complication that the test criterion may not be distribution-free under the null hypothesis.

Tests may also be constructed using other "distances" than those mentioned in Section 2 and above (see also [5], pp. 148–149, and [6], p. 10 in this connection), involving other "estimators" than those of Sections 2 and 3, and involving other modifications of the notion of "distance methods" as motivated in this section.

2. Testing normality. For convenience we divide this section into several subsections.

2.1. The computation of $\delta(G_n^*, N^{**})$ offers considerable difficulty; unpublished work on this subject has been done by Blackman. The distribution of $\delta(G_n^*, N^{**})$ under the null hypothesis is still unknown. (It is easy to verify that, when $G(x)$ is a member of N^{**}, the distribution of $\delta(G_n^*, N^{**})$ does not depend on $G(x)$. Thus the composite null hypothesis determines uniquely the distribution of the test criterion.) In Section 4 we shall give an example of another problem of testing hypotheses where the limiting distribution of the minimum distance criterion is explicitly calculated. In the present Section 2 we shall consider some other "distance" tests. (For the case where either the mean or variance is assumed known, the test corresponding to that discussed in this section is being studied by Darling [18]; the suggestion of using such tests is apparently due to Cramér.) Let

$$\bar{x} = \frac{1}{n} \sum_1^n x_i, \qquad s^2 = \frac{1}{n} \sum_1^n x_i^2 - \bar{x}^2.$$

Let $N(x \mid \bar{x}, s^2)$ be the normal distribution function with mean $\bar{x}$ and variance s^2. One can base the test of normality on $v_n = \delta(G_n^*(x), N(x \mid \bar{x}, s^2))$, with the large values critical. It is easy to see that, when $G(x)$ is actually a member of

N^{**}, the distribution of v_n does not depend on $G(x)$. The distribution of v_n does not seem easy to obtain, except, for example, by Monte Carlo methods. Another test criterion, similar to the above but of the "ω^2" type, is to base the test on

$$w_n = \frac{1}{n}\int_0^1 Z_{r,n}^2\, dr$$

where $Z_{r,n}$ is defined in 2.2.[3] (The idea of the "ω^2" test, which is defined precisely in Section 6, is due to von Mises, with a modification by Smirnov.) Sections 2.2 to 2.6 are concerned with the limiting distribution of nw_n as $n \to \infty$ when G is normal. The power of the tests based on v_n and w_n is discussed in Section 5. In Section 3 we treat briefly an example which illustrates the construction of test criteria which are similar to v_n and w_n but which may use "inefficient" statistics to estimate which member of N^{**} G is (if it is); such techniques may have obvious practical importance.

2.2. Let $x_1, x_2, \cdots$ be independent, identically distributed Gaussian random variables with zero mean and unit variance. Let $G_n^*(x)$ be as defined in Section 1. We shall use the following notation:

$$\phi(x) = \phi(-x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}, \qquad I(x) = \int_x^\infty \phi(y)\,dy,$$

$$J(y) = I^{-1}(1-y) = \{x \mid y = 1 - I(x)\}, \qquad g(x) = \phi(x)/I(x).$$

We also let $[z]$ = smallest integer $\geqq z$. For $0 \leqq r \leqq 1$, we put $U_{r,n} = [rn]$th from the bottom among the ordered $x_1, \cdots, x_n$. Finally, let[4]

$$Z_{r,n} = \sqrt{n}[G_n^*(\sqrt{S_n'}J(r) + \bar{x}_n) - r]$$

where

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^n x_i, \qquad S_n = \frac{1}{n}\sum_{i=1}^n x_i^2, \qquad S_n' = S_n - \bar{x}_n^2.$$

(Remark: S_n and S_n' may be used equally well in the definition of $Z_{r,n}$ for the purpose of obtaining the limiting distribution; in applications S_n' should probably be used.)

2.3. We shall show that for $0 \leqq r_1 < r_2 < \cdots < r_k \leqq 1$, the quantities $Z_{r_i,n}$, where $1 \leqq i \leqq k$, are asymptotically jointly normally distributed with zero means and covariance function (for $Z_{s,n}$ and $Z_{t,n}$ as $n \to \infty$)

$$K(s, t) = \min(s, t) - st - (2\pi)^{-1}(1 + J(s)J(t)/2)e^{-[J(s)^2+J(t)^2]/2}. \tag{2.1}$$

We shall show here that (2.1) follows from the fact proved in Section 2.4, that $\sqrt{n}\bar{x}_n$, $\sqrt{n}(S_n - 1)$, and $\sqrt{n}(U_{r_i,n} - J(r_i))$, where $1 \leqq i \leqq k$, are jointly

[3] This definition is equivalent to that given in the summary at the beginning of the paper.

[4] Elsewhere in this paper, in the summary, for example, the conventional symbol s^2 is sometimes used instead of the typographically easier (here) symbol S_n'. Both represent the same thing. Also $\bar{x}_n$ and $\bar{x}$ are used interchangeably in a manner to cause no confusion.

asymptotically normal with means 0 and covariance matrix (presented in the same order) given by

$$
(2.2)\qquad \begin{pmatrix} 1 & 0 & 1 & 1 & \cdots & 1 \\ 0 & 2 & J(r_1) & J(r_2) & \cdots & J(r_k) \\ 1 & J(r_1) & \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1k} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & J(r_k) & \lambda_{k1} & \lambda_{k2} & \cdots & \lambda_{kk} \end{pmatrix},
$$

$$
\lambda_{ij} = \lambda_{ji} = \frac{r_i(1 - r_j)}{\phi(J(r_i))\phi(J(r_j))}, \qquad r_i \leqq r_j .
$$

In fact, all these results are well known ([1], p. 364, 369) except for the joint normality and the last k entries in the first two rows (and columns).

Assume then for the moment that the above is proved. The event $\{Z_{r,n} \leqq z\}$ may be written as $\{[\text{number of } x_i \text{ which are } < \sqrt{S'_n}J(r) + \bar{x}_n] \leqq nr + \sqrt{n}z\}$. This in turn is the same as $\{U_{r+z/\sqrt{n},\, n} \geqq \sqrt{S'_n}J(r) + \bar{x}_n\}$, or

$$
\{\sqrt{n}\bar{x}_n + \sqrt{n}J(r)(\sqrt{S'_n} - 1) - \sqrt{n}(U_{r+z/\sqrt{n},\,n} - J(r + z/\sqrt{n}))
$$
$$
\leqq \sqrt{n}[J(r + z/\sqrt{n}) - J(r)]\}.
$$

As $n \to \infty$, neglecting terms of higher order in probability, we may replace $\sqrt{n}[J(r + z/\sqrt{n}) - J(r)]$ by $zJ'(r)$ and $\sqrt{n}(\sqrt{S'_n} - 1)$ by $\frac{1}{2}\sqrt{n}(S_n - 1)$. We conclude that

$$
\lim_{n\to\infty} P\{Z_{r_1,n} \leqq z_1, \cdots, Z_{r_k,n} \leqq z_k\}
$$

$$
(2.3)\qquad = \lim_{n\to\infty} P\left\{\frac{1}{J'(r_i)}[\sqrt{n}\bar{x}_n + \tfrac{1}{2}\sqrt{n}J(r_i)(S_n - 1) - \sqrt{n}(U_{r_i+z_i/\sqrt{n},\,n} - J(r_i + z_i/\sqrt{n}))] \leqq z_i, \quad i = 1, \cdots, k\right\}.
$$

Thus, $Z_{r_1,n}, Z_{r_2,n}, \cdots, Z_{r_k,n}$ are jointly asymptotically normal, with the same limiting distribution as the quantities

$$
[\sqrt{n}\bar{x}_n + \tfrac{1}{2}\sqrt{n}J(r_i)(S_n - 1) - \sqrt{n}(U_{r_i,n} - J(r_i))]/J'(r_i), \quad i = 1, \cdots, k.
$$

From this and (2.2), the result (2.1) follows by direct computation and the fact that $J'(r) = 1/\phi(J(r))$.

2.4. It remains to prove the joint asymptotic normality of the quantities mentioned in the paragraph following (2.1), and to verify the last k entries of the first two rows (and columns) of (2.2). We shall in fact compute here only the limiting distribution of $\sqrt{n}\bar{x}_n$, $\sqrt{n}(S_n - 1)$, and $\sqrt{n}(U_{r,n} - J(r))$ where $0 < r < 1$; from this will follow the desired result of (2.2), and the method of proof of joint asymptotic normality for the $k + 2$ random variables previously

mentioned will be evident from that for the case $k = 1$ considered here (note especially p. 369 of [1] in this connection).

We begin by introducing a process which will enable us to simplify this computation. For fixed n, let $Y_{[rn]}$ be a chance variable whose distribution function is that of $U_{r,n}$ above. Let Y_i, where $1 \leqq i \leqq n$ and $i \neq [rn]$, be random variables whose joint conditional distribution, given that $Y_{[rn]} = y$, is such that these $n - 1$ random variables are (conditionally) independent with Y_i having a (conditional) truncated normal distribution given by

$$P\{Y_i < z \mid Y_{[rn]} = y\} = \begin{cases} \min[1, & (1 - \mathrm{I}(z))/(1 - \mathrm{I}(y))],\ 1 \leqq i < [rn] \\ \max[0, & 1 - \mathrm{I}(z)/\mathrm{I}(y)],\ [rn] < i \leqq n. \end{cases}$$

If now $Y_1, \cdots, Y_n$ are reordered with probability $1/n!$ for each possible reordering, and the resulting reordered variables are labeled $X_1^1, \cdots, X_n^1$, it is easily verified that $X_1^1, \cdots, X_n^1$ have the same joint distribution as the $x_1, \cdots, x_n$ considered in Section 1. We shall use the process $Y_1, \cdots, Y_n$ to compute the conditional distribution of S_n and $\bar{x}_n$, given that $U_{r,n} = y = J(r) + w/\sqrt{n}$, say. Let

$$([nr] - 1)\bar{X}_n^1 = \sum_1^{[rn]-1} Y_i, \qquad (n - [nr])\bar{X}_n^2 = \sum_{[rn]+1}^{n} Y_i,$$

$$([nr] - 1)S_n^1 = \sum_1^{[rn]-1} Y_i^2, \qquad (n - [nr])S_n^2 = \sum_{[rn]+1}^{n} Y_i^2.$$

It is easy to compute that for the truncated normal distribution from y to ∞ mentioned above, the first four moments about the origin are

$$\mu_1(y) = g(y), \qquad \mu_2(y) = 1 + y\,g(y),$$

$$\mu_3(y) = (y^2 + 2)\,g(y), \qquad \mu_4(y) = 3 + (y^3 + 3y)\,g(y).$$

The corresponding moments for the truncated distribution from $-\infty$ to y are clearly $-\mu_1(-y)$, $\mu_2(-y)$, $-\mu_3(-y)$, and $\mu_4(-y)$. From these and [1], p. 364, we obtain that

$$A_n = \sqrt{[nr] - 1}\,(\bar{X}_n^1 + g(-y)), \qquad B_n = \sqrt{[nr] - 1}(S_n^1 - 1 + y\,g(-y))$$

are asymptotically conditionally normal with means 0 and covariance matrix

$$\begin{pmatrix} 1 - y\,g(-y) - [g(-y)]^2 & -(y^2 + 1)\,g(-y) - y[g(-y)]^2 \\ -(y^2 + 1)\,g(-y) - y[g(-y)]^2 & 2 - (y^3 + y)\,g(-y) - y^2[g(-y)]^2 \end{pmatrix},$$

and that $C_n = \sqrt{n - [nr]}\,(\bar{X}_n^2 - g(y))$ and $D_n = \sqrt{n - [nr]}\,(S_n^2 - 1 - yg(y))$ are asymptotically conditionally normal with means 0 and covariance matrix

$$\begin{pmatrix} 1 + y\,g(y) - [g(y)]^2 & (y^2 + 1)\,g(y) - y[g(y)]^2 \\ (y^2 + 1)\,g(y) - y[g(y)]^2 & 2 + (y^3 + y)\,g(y) - y^2[g(y)]^2 \end{pmatrix},$$

with (A_n, B_n) conditionally independent of (C_n, D_n). By using the Liapounoff

condition one proves that the approach to normality is uniform for any finite w-interval. Let $O_p(\ )$ denote "order of ( ) in probability" (e.g., [9]). Noting that $\sqrt{[nr]} - \sqrt{nr} = O(1/\sqrt{n})$ and that $\sqrt{n}(y/n) = O_p(1/\sqrt{n})$, we obtain as an expression for $\sqrt{n}\bar{x}_n$, given that $U_{r,n} = y$,

$$\sqrt{n}\bar{x}_n = \sqrt{r}A_n + \sqrt{1-r}\,C_n + \sqrt{n}[-r\,g(-y) + (1-r)\,g(y)] + O_p(1/\sqrt{n}).$$

But

$$-r\,g\,(-J(r) - w/\sqrt{n}) + (1-r)\,g\,(J(r) + w/\sqrt{n})$$
$$= \frac{w}{\sqrt{n}} \cdot \frac{[\phi(J(r))]^2}{r(1-r)} + O_p\left(\frac{1}{n}\right),$$

so that

$$\sqrt{n}\bar{x}_n = \sqrt{r}A_n + \sqrt{1-r}\,C_n + w\,\frac{[\phi(J(r))]^2}{r(1-r)} + O_p\left(\frac{1}{\sqrt{n}}\right).$$

In similar fashion, we obtain

$$\sqrt{n}(S_n - 1) = \sqrt{r}B_n + \sqrt{1-r}D_n + wJ(r)\,\frac{[\phi(J(r))]^2}{r(1-r)} + O_p\left(\frac{1}{\sqrt{n}}\right).$$

In both cases the $O_p(1/\sqrt{n})$ term is uniform in every bounded interval of w. Since $W_n = \sqrt{n}(U_{r,n} - J(r))$ is asymptotically normal with mean 0 and variance $r(1-r)/[\phi(J(r))]^2$, the desired asymptotic joint normality follows easily from the last two displayed expressions. The covariance of $\sqrt{n}\bar{x}_n$ and W_n is most easily computed as $E\{W_n \cdot E\{\sqrt{n}\bar{x}_n \mid W_n\}\}$, that of $\sqrt{n}(S_n - 1)$ and W_n being computed similarly; these give the desired results for the last k entries of the first two rows (and columns) of (2.2).

2.5. We now show that there exists a representation of any Gaussian process with mean identically zero and a covariance function like that of (2.1), whose sample functions are continuous with probability one (w.p. 1).

Let $W(t)$, $0 \leqq t \leqq 1$, be the Kac-Siegert representation ([12]) of a Gaussian process with continuous covariance function $K'(s, t)$. Let $\{\lambda_k\}$ be the eigenvalues and $\{\varphi_k(t)\}$ the corresponding normalized eigenfunctions of $K'(s, t)$. Suppose furthermore that $g(t)$ in L^2 is such that

$$K'(s, t) - g(s)g(t) = K''(s, t)$$

is a covariance function. We shall show how to construct explicitly a process $Z(t)$ with covariance function $K''(s, t)$ in terms of the process $W(t)$. (All processes studied in this section are to have mean zero.)

We first prove two lemmas.

LEMMA 1. *A necessary and sufficient condition that $K''(s, t)$ be positive definite is that*

$$\beta^2 = \sum_{k=1}^{\infty} \frac{g_k^2}{\lambda_k} \leqq 1, \qquad g_k = \int_0^1 g(t)\,\varphi_k(t)\,dt.$$

(If a λ_k is 0, it follows immediately from the positivity of $K''(s, t)$ that $g(t)$ is orthogonal to all the eigenfunctions belonging to λ_k; thus the corresponding g_k vanish and we interpret g_k^2/λ_k as zero.)

Only the necessity of the condition is used in the application of the lemma. The proof of sufficiency is included because it is so brief and sufficiency seems to be of interest.

1. *Necessity*. Set $\psi(t) = \sum_1^n v_k \varphi_k(t)$ and note that by definition

$$\int_0^1 \int_0^1 (K'(s, t) - g(s)\, g(t))\, \psi(s)\, \psi(t)\, ds\, dt \geqq 0.$$

Evaluating the double integral we obtain

$$\sum_1^n v_k^2 \lambda_k \geqq \left(\sum_1^n v_k g_k\right)^2,$$

and setting $v_k = g_k/\lambda_k$

$$\sum_1^n \frac{g_k^2}{\lambda_k} \geqq \left(\sum_1^n \frac{g_k^2}{\lambda_k}\right)^2.$$

Thus $\sum_1^n g_k^2/\lambda_k \leqq 1$ for *every* n and the theorem follows.

2. *Sufficiency*. If $\sum_1^\infty g_k^2/\lambda_k \leqq 1$ we have

$$\left(\sum_1^\infty v_k g_k\right)^2 = \left(\sum_1^n v_k \sqrt{\lambda_k}\, \frac{g_k}{\sqrt{\lambda_k}}\right)^2 \leqq \left(\sum_1^\infty v_k^2 \lambda_k\right)\left(\sum_1^\infty \frac{g_k^2}{\lambda_k}\right) \leqq \sum_1^\infty v_k^2 \lambda_k,$$

and positiveness of $K'(s, t) - g(s)\, g(t)$ follows.

LEMMA 2. *The series* $\sum_1^\infty g_k \varphi_k(t)$ *converges uniformly (to* $g(t)$*).*

We have $\left|\sum_m^n g_k \varphi_k(t)\right| \leqq \sqrt{\sum_m^n g_k^2/\lambda_k}\, \sqrt{\sum_m^n \lambda_k \varphi_k^2(t)}$. By Mercer's theorem, $\sum_1^\infty \lambda_k \varphi_k^2(t) = K(t, t)$. Hence $\left|\sum_m^n g_k \varphi_k(t)\right| \leqq \sqrt{K(t, t)}\, \sqrt{\sum_m^n g_k^2/\lambda_k}$. Since $\sum g_k^2/\lambda_k$ converges (by Lemma 1), Lemma 2 follows.

We are now ready to prove that

$$Z(t) = W(t) - \frac{(1 - \sqrt{1 - \beta^2})}{\beta^2}\, g(t) \sum_{k=1}^\infty \frac{g_k}{\lambda_k} \int_0^1 W(t)\, \varphi_k(t)\, dt \tag{2.4}$$

has the covariance function $K''(s, t)$.

Note first that the chance variables

$$\left\{\frac{1}{\sqrt{\lambda_k}} \int_0^1 W(t) \varphi_k(t)\; dt\right\}$$

are independently and normally distributed, with mean 0 and variance 1. Since $\sum_1^\infty g_k^2/\lambda_k < \infty$, the series

$$\sum_1^\infty \frac{g_k}{\sqrt{\lambda_k}} \cdot \frac{1}{\sqrt{\lambda_k}} \int_0^1 W(t) \varphi_k(t)\; dt$$

converges in the mean (and even w.p. 1) and thus defines a random variable.

To calculate the covariance of $Z(t)$ we need

$$E\left\{W(t)\sum_{k=1}^{\infty}\frac{g_k}{\lambda_k}\int_0^1 W(t)\varphi_k(t)\,dt\right\}.$$

But $E\{W(t)\int_0^1 W(t)\varphi_k(t)\,dt\} = \int_0^1 E\{W(t)W(t)\}\varphi_k(t)\,dt = \int_0^1 K'(t, t)\varphi_k(t)\,dt = \lambda_k\varphi_k(t)$, and hence

$$E\left\{W(t)\sum_{k=1}^{\infty}\frac{g_k}{\lambda_k}\int_0^1 W(t)\varphi_k(t)\,dt\right\} = \sum_{k=1}^{\infty} g_k\varphi_k(t) = g(t).$$

The covariance function of $Z(t)$ now comes out to be

$$K'(s,t) - 2\beta^{-2}(1-\sqrt{1-\beta^2})\,g(s)g(t) + \beta^{-2}(1-\sqrt{1-\beta^2})^2\,g(s)g(t)$$
$$= K'(s,t) - g(s)g(t) = K''(s,t).$$

Since $g(t)$ is continuous (by Lemma 2) it follows, in particular, that the sample functions of $Z(t)$ are continuous w.p. 1 if the sample functions of $W(t)$ are continuous w.p. 1. Now let $K'(s, t) = \min(s, t) - st$. Then the sample functions of $W(t)$ are continuous w.p. 1. The application of the above result twice (once for each remaining negative product in (2.1)) then proves that the representation $Z(t)$, which is Gaussian with mean zero and covariance function (2.1), has continuous sample functions, w.p. 1.

2.6. From the results of Sections 2.2 to 2.5, it is easily verified that the demonstration given by Kac on pp. 197–198 of [10] for the case $K(s, t) = \min(s, t) - st$ carries over with only slight modifications to the case now under discussion where $K(s, t)$ is given by (2.1). We conclude that

$$\lim_{n\to\infty} P\{nw_n < a\} = P\{W < a\}, \tag{2.5}$$

where $W = \int_0^1 [Z(t)]^2\,dt$ and $Z(t)$ is Gaussian with covariance function given by (2.1) and sample functions continuous w.p. 1. Modifying slightly the technique of Kac and Siegert [11], [12] as applied on pp. 199–200 of [10], we now study the distribution of Z. We write

$$K(s,t) = K^*(s,t) - h_1(s)\,h_1(t) - h_2(s)\,h_2(t) \qquad 0 \le s, t \le 1, \tag{2.6}$$

where $K^*(s, t) = \min(s, t) - st$ and $h_k(s) = (2\pi)^{-1/2}[J(s)/\sqrt{2}]^{k-1}e^{-[J(s)]^2/2}$ for $k = 1, 2$. Let $\lambda_1, \lambda_2, \cdots$ be the eigenvalues (all positive, since K is positive definite) of the integral equation

$$\int_0^1 K(x,y)\,\varphi(y)\,dy = \lambda\,\varphi(x). \tag{2.7}$$

Following the demonstration of [10], we conclude that the characteristic function of W is

$$Ee^{iW\xi} = \prod_{j=1}^{\infty}(1-2i\xi\lambda_j)^{-1/2}. \tag{2.8}$$

Thus, we may express W as

$$W = \sum_{j=1}^{\infty} R_j, \tag{2.9}$$

where the R_j are independent and R_j has density function

$$\frac{1}{\sqrt{2\pi r \lambda_j}} e^{-r/2\lambda_j}, \qquad r > 0. \tag{2.10}$$

(We remark that it may of course be easier to obtain the convolution with itself of the distribution of W rather than the latter itself. This could be used if one based a test of normality on the sum of two W_m's, each computed from half the sample of size $n = 2m$.)

We now give a procedure for finding the λ_j. For any eigenvalue λ and corresponding eigenfunction φ (not necessarily normalized) of (2.7), write

$$C_i = \int_0^1 h_i(y)\, \phi(y)\, dy, \qquad i = 1, 2. \tag{2.11}$$

We can rewrite (2.7) as

$$\int_0^1 K^*(x, y)\, \varphi(y)\, dy - C_1 h_1(x) - C_2 h_2(x) = \lambda\, \varphi(x). \tag{2.12}$$

Differentiating twice with respect to x and writing $\mu^2 = 1/\lambda$ (we may consider $\mu > 0$ in the sequel), we obtain

$$\varphi''(x) + \mu^2 \varphi(x) = -C_1 \mu^2 h_1''(x) - C_2 \mu^2 h_2''(x). \tag{2.13}$$

Any eigenvalue λ and eigenfunction $\phi(x)$ of (2.7) satisfy (2.13), (2.11), and

$$\varphi(0) = \phi(1) = 0. \tag{2.14}$$

Conversely if λ and $\phi(x)$ satisfy these conditions they are an eigenvalue and eigenfunction of (2.7). For let

$$\int_0^1 K(x, y)\, \phi(y)\, dy = \lambda\, \theta(x).$$

As we obtained (2.13) we get

$$\theta''(x) + \mu^2 \phi(x) = -C_1 \mu^2 h_1''(x) - C_2 \mu^2 h_2''(x).$$

Hence $\theta''(x) = \phi''(x)$, and since also $\theta(0) = \theta(1) = 0$ we have $\theta(x) = \phi(x)$.

Our problem now is to find a value μ for which there exist a function $\phi(x)$ and constants C_1 and C_2 satisfying (2.13), (2.14), and (2.11). For given C_1 and C_2 the general solution of (2.13) can be written

$$\phi(x) = A \sin \mu x + B \cos \mu x - \mu C_1 g_1(x) - \mu C_2 g_2(x), \tag{2.15}$$

where

$$g_i(x) = \int_x^{1/2} h_i''(t) \sin \mu(t - x)\, dt. \tag{2.16}$$

Applying conditions (2.14) and (2.11) gives

$$0 = B - \mu C_1 g_1(0) - \mu C_2 g_2(0), \qquad 0 = A \sin \mu + B \cos \mu - \mu C_1 g_1(1) - \mu C_2 g_2(1),$$

$$\begin{aligned} C_i = A \int_0^1 h_i(x) \sin \mu x \, dx + B \int_0^1 h_i(x) \cos \mu x \, dx \\ -\mu C_1 \int_0^1 h_i(x)\, g_1(x)\, dx - \mu C_2 \int_0^1 h_i(x)\, g_2(x)\, dx. \end{aligned} \tag{2.17}$$

These equations have a nontrivial solution for A, B, C_1, and C_2 if and only if the determinant $D(\mu)$ of the coefficient of these four quantities is zero. Hence the eigenvalues of (2.7) are determined by the roots of $D(\mu) = 0$.

The following method of computing $D(\mu)$ is due to R. J. Walker, to whom the authors are greatly obliged.

We first note some pertinent properties of $h_i(x)$ and $g_i(x)$.

$$h_1(1 - x) = h_1(x), \qquad h_1(0) = 0, \qquad h_1(\tfrac{1}{2}) = 1/\sqrt{2\pi}, \qquad h_1'(\tfrac{1}{2}) = 0;$$

$$h_2(1 - x) = -h_2(x), \qquad h_2(0) = 0, \qquad h_2(\tfrac{1}{2}) = 0, \qquad h_2'(\tfrac{1}{2}) = 1/\sqrt{2};$$

$$\begin{aligned} g_1(1 - x) &= \int_{1-x}^{1/2} h_1''(t) \sin \mu(t - 1 + x)\, dt = \int_x^{1/2} h_1''(1 - s) \sin \mu(s - x)\, ds \\ &= g_1(x); \end{aligned}$$

and similarly

$$g_2(1 - x) = -g_2(x).$$

It follows that $g_1(1) = g_1(0)$ and $g_2(1) = -g_2(0)$, and

$$\int_0^1 h_i(x)\, g_j(x)\, dx = \begin{cases} 2 \int_0^{1/2} h_i(x) g_i(x), & j = i; \\ 0, & j \neq i. \end{cases}$$

Also, using $\sin \mu x = \sin \frac{1}{2}\mu \cos \mu(\frac{1}{2} - x) - \cos \frac{1}{2}\mu \sin \mu(\frac{1}{2} - x)$, we get

$$\int_0^1 h_1(x) \sin \mu x \, dx = 2 \sin \tfrac{1}{2}\mu \int_0^{1/2} h_1(x) \cos \mu(\tfrac{1}{2} - x)\, dx,$$

$$\int_0^1 h_2(x) \sin \mu x \, dx = -2 \cos \tfrac{1}{2}\mu \int_0^{1/2} h_2(x) \sin \mu(\tfrac{1}{2} - x)\, dx,$$

with similar reductions for the coefficients of B in (2.17).

Introducing these simplifications we get by direct computation $D(\mu) = -2D_1(\mu)\, D_2(\mu)$, where

$$\begin{aligned} D_1(\mu) = \cos \tfrac{1}{2}\mu \left[1 + 2\mu \int_0^{1/2} h_1(x)\, g_1(x)\, dx \right] \\ -2\mu g_1(0) \int_0^{1/2} h_1(x) \cos \mu(\tfrac{1}{2} - x)\, dx, \end{aligned} \tag{2.18}$$

$$D_2(\mu) = \sin \tfrac{1}{2}\mu \left[1 + 2\mu \int_0^{1/2} h_2(x)\, g_2(x)\, dx \right] \tag{2.19}$$
$$\qquad - 2\mu g_2(0) \int_0^{1/2} h_2(x) \sin \mu(\tfrac{1}{2} - x)\, dx.$$

These equations can be put in a form more suitable for computation. Integrating (2.16) by parts gives, for $i = 1$,

$$g_1(x) = -\frac{\mu}{\sqrt{2\pi}} \cos \mu(\tfrac{1}{2} - x) + \mu h_1(x) - \mu^2 \int_x^{1/2} h_1(t) \sin \mu(t - x)\, dt$$

$$g_1(0) = -\frac{\mu}{\sqrt{2\pi}} \cos \tfrac{1}{2}\mu - \mu^2 \int_0^{1/2} h_1(t) \sin \mu t\, dt.$$

Putting these in (2.18) gives

$$D_1(\mu) = \cos \tfrac{1}{2}\mu \left[1 + 2\mu^2 \int_0^{1/2} h_1^2(x)\, dx \right]$$
$$+ 2\mu^3 \left[\iint_{S+T} h_1(x)\, h_1(t) \cos \mu(\tfrac{1}{2} - x) \sin \mu t\, dA \right.$$
$$\left. - \iint_T h_1(x)\, h_1(t) \sin \mu(t - x) \cos \tfrac{1}{2}\mu\, dA \right],$$

where S is the triangle bounded by the lines $t = 0$, $t = x$, and $x = \frac{1}{2}$, and T the triangle bounded by $x = 0$, $x = t$, and $t = \frac{1}{2}$. The sum of the integrals over T reduces to

$$\iint_T h_1(x)\, h_1(t) \cos \mu(\tfrac{1}{2} - t) \sin \mu x\, dA,$$

which equals the integral over S. Hence, finally

$$D_1(\mu) = \cos \tfrac{1}{2}\mu \left[1 + 2\mu^2 \int_0^{1/2} h_1^2(x)\, dx \right] \tag{2.20}$$
$$+ 4\mu^3 \int_0^{1/2} h_1(x) \sin \mu x\, dx \int_x^{1/2} h_1(t) \cos \mu(\tfrac{1}{2} - t)\, dt.$$

Similarly, (2.19) becomes

$$D_2(\mu) = \sin \tfrac{1}{2}\mu \left[1 + 2\mu^2 \int_0^{1/2} h_2^2(x)\, dx \right] \tag{2.21}$$
$$+ 4\mu^3 \int_0^{1/2} h_2(x) \sin \mu x\, dx \int_x^{1/2} h_2(t) \sin \mu(\tfrac{1}{2} - t)\, dt.$$

The method just developed for obtaining the eigenvalues of (2.7) seems more accurate and computationally simpler than other methods, such as those employing trial functions. In Section 6 the smallest few zeros of the functions $D_1(\mu)$ and $D_2(\mu)$ of (2.20) and (2.21) are tabulated, and an approximation for the

distribution of W is thereby obtained. This approximation is compared with empirical distribution functions of nw_n obtained by sampling. Empirical distribution functions of nw_n and $\sqrt{nw_n}$ are tabulated, and certain other interesting results of the sampling experiments are noted (for example, the joint distribution of the classical Kolmogoroff and von Mises statistics D_n and ω_n^2, which are defined precisely in Section 6).

3. Tests using quantiles. The relative ease of computing the limiting distributions of various possible test criteria of the type considered in this paper will of course depend on the particular problem. Thus, the use of the sample mean and variance in non-normal cases may lead to more complicated results than those of Section 2. A tool which may be used in all cases (where the hypothesized family has finitely many natural parameters, is normal or not, where a simple sufficient statistic does or does not exist, etc.) of density functions, with about equal complexity in all cases, is the use of sample quantiles (as many as necessary) to estimate the "true" d.f. if the null hypothesis is true. (Of course, the more unknown parameters and hence quantiles which must be used, the messier will be the result.) For computational reasons, tests constructed in this manner will sometimes be more practical to use than those involving a sufficient statistic. The results on power in Section 5 apply also here.

As an example, suppose the pth sample quantile $U_{p,n}$, $0 < p < 1$, is to be used to estimate the corresponding population parameter of a family of d.f.'s $F(x - \theta)$ for $-\infty < \theta < \infty$, in testing whether or not the "true" d.f. is a member of this family. We suppose without loss of generality that $F(0) = p$, and we denote by f the density function of F. We assume f to be continuous and positive in a neighborhood of 0.

Then letting $\psi = F^{-1}$ and $Z_{r,n} = \sqrt{n}[F_n(\psi(r) + U_{p,n}) - r]$, an argument like that of Sections 2.3 and 2.4 leads easily to the conclusion that, for $0 \leqq r_1 \leqq \cdots \leqq r_k \leqq 1$, the limiting distribution of $Z_{r_i,n}$, $1 \leqq i \leqq k$, is the same as that of

$$\sqrt{n}[U_{p,n} - [U_{r_i,n} - \psi(r_i)]] \,/\, \psi'(r_i), \qquad 1 \leqq i \leqq k,$$

and, in particular, is Gaussian. Putting $\gamma(r) = f(\psi(r))$ and

$$g(r) = \sqrt{p(1-p)}\,\frac{\gamma(r)}{\gamma(0)} - \frac{\min[p, r] - pr}{\sqrt{p(1-p)}},$$

we obtain for the limiting covariance function, for $0 \leqq s,t \leqq 1$,

$$\begin{aligned} K(s, t) &= \gamma(s)\,\gamma(t)\left[\frac{p(1-p)}{[\gamma(0)]^2} - \frac{\min(p, s) - ps}{\gamma(0)\,\gamma(s)}\right. \\ &\qquad\qquad \left. - \frac{\min(p, t) - pt}{\gamma(0)\,\gamma(t)} + \frac{\min(s, t) - st}{\gamma(s)\,\gamma(t)}\right] \\ &= g(s)\,g(t) - \frac{[\min(p, s) - ps][\min(p, t) - pt]}{p(1-p)} + \min(s, t) - st \end{aligned}$$

$$
= g(s)\,g(t) + \begin{cases} \min(s, t) - st/p & s, t \leqq p, \\ \min(s - p, t - p) - \dfrac{(s-p)(t-p)}{1-p} & s, t \geqq p, \\ 0 & \text{otherwise.} \end{cases}
$$

For computing purposes (e.g., by Monte Carlo methods), the last form of the covariance function is useful. Let $X_1(t)$ and $X_2(t)$ be Gaussian, each with the familiar covariance function $\min(s, t) - st$, and let X be a normal random variable with mean zero and variance 1, where X, X_1, and X_2 are independent. Defining $X_1(t) = X_2(t) = 0$ if $t < 0$ or $t > 1$, let

$$
Z(t) = g(t)X + \sqrt{p}\, X_1\!\left(\frac{t}{p}\right) + \sqrt{1-p}\, X_2\!\left(\frac{t-p}{1-p}\right), \quad 0 \leqq t \leqq 1.
$$

Then Z has the covariance function $K(s, t)$. As an example, if F is normal with unit variance and $p = \frac{1}{2}$, we have

$$
(Zt) = [\tfrac{1}{2}e^{-\psi(t)^2/2} = \min(t, 1 - t)]X + \frac{1}{\sqrt{2}}\, X_1(2t) + \frac{1}{\sqrt{2}}\, X_2(2t - 1).
$$

4. The rectangular distribution. In Section 4.1 we give an example where the minimum distance method statistic can be computed explicitly; in Section 4.2 we comment on the limiting distribution of test criteria of the type treated in Section 2 in the present case.

4.1. Let

$$
F(x; \theta) = \begin{cases} 0 & x < \theta - \frac{1}{2}, \\ x - \theta + \frac{1}{2} & \theta - \frac{1}{2} \leqq x \leqq \theta + \frac{1}{2}, \\ 1 & \theta + \frac{1}{2} < x, \end{cases}
$$

and let R denote the family of all such distribution functions for $-\infty < \theta < \infty$. It is desired to test the hypothesis that $x_1, x_2, \cdots, x_n$ are independently and identically distributed according to some member of R. We will be concerned with computations when this hypothesis is true (see Section 5 for remarks on power which apply also here), and denote by G the true member of R (i.e., the distribution function of X_1). G_n^* is as defined in Section 1. Let

$$
D_n^+ = \sup_x (G_n^*(x) - G(x)), \qquad D_n^- = \sup_x (G(x) - G_n^*(x)).
$$

In the present example the minimum distance criterion is easily seen to be

$$
\delta(G_n^*, R) = \inf_\theta \sup_x |G_n^*(x) - F(x; \theta)| = \tfrac{1}{2}(D_n^+ + D_n^-). \tag{4.1}
$$

The joint limiting distribution of $\sqrt{n}D_n^+$ and $\sqrt{n}D_n^-$ (as $n \to \infty$) is given by Doob ([13], p. 403) for $x,y > 0$ as

$$
\begin{aligned}
&\lim_{n\to\infty} P\{\sqrt{n}D_n^- \leqq x, \sqrt{n}D_n^+ \leqq y\} \\
&\qquad = 1 - \sum_{m=1}^{\infty} \{e^{-2(mx+(m-1)y)^2} + e^{-2(my+(m-1)x)^2} - 2e^{-2m^2(x+y)^2}\} \qquad (4.2) \\
&\qquad = G(x, y), \quad \text{say.}
\end{aligned}
$$

Let $g(x, y) = \partial^2 G(x, y)/\partial x \partial y$. The series in (4.2) is uniformly and absolutely convergent and differentiable with respect to x and y outside of any neighborhood of the origin. Using the fact that the mixed derivative of the expression $e^{-2(ax+by)^2}$ is $4ab(u^2 - 1)e^{-u^2/2}$ where $u = (2ax + 2by)^2$, we obtain for the mth term of the series for $\int_0^z g(y, z - y)\,dy$ the expression

$$
\begin{gathered}
4m(m-1)[H(2mz) - H(2(m-1)z)] + 4m(4m^2z^2 - 1)H(2mz), \\
H(x) = \int_x^\infty (u^2 - 1)e^{-u^2/2}\,du = xe^{-x^2/2}.
\end{gathered} \tag{4.3}
$$

The density function corresponding to the limiting distribution function of $B_n = \sqrt{n}(D_n^+ + D_n^-)$ is a sum of expressions given in (4.3). For $\sqrt{n}\delta(G_n^*, R) = \frac{1}{2}B_n = U$, say, the expression corresponding to (4.3) is

$$
8m(m-1)[H(4mu) - H(4(m-1)u)] + 8m(16m^2u^2 - 1)H(4mu). \tag{4.4}
$$

The series is absolutely convergent and in the sum of (4.4) from 1 to ∞, the coefficient of the expression $H(4ku)$ for $k \geqq 0$ is $8k(16k^2u^2 - 3)$, so that the density function of the limiting cumulative distribution function of U is

$$
32u \sum_{m=1}^{\infty} m^2(16m^2u^2 - 3)e^{-8m^2u^2}. \tag{4.5}
$$

Outside any given neighborhood of the origin, all terms of (4.5) are positive except for a finite number. Thus, integrating (4.5) we obtain for $u > 0$

$$
\lim_{n\to\infty} P\{\sqrt{n}\delta(G_n^*, R) \leqq u\} = 1 - \sum_{m=1}^{\infty} (32m^2u^2 - 2)e^{-8m^2u^2}. \tag{4.6}
$$

It is also of interest that in this example we can compute the limiting distribution of the minimum distance *estimator* of θ, namely, the random variable $T_n = T_n(x_1, \cdots, x_n)$ defined by

$$
\delta(G_n^*(y), F(y; T_n)) = \delta(G_n^*, R), \tag{4.7}
$$

which is satisfied by $T_n - \theta = \frac{1}{2}(D_n^- - D_n^+)$ when θ is the true parameter value. An analysis similar to that given above shows that, for $t > 0$,

$$
\lim_{n\to\infty} P\{\sqrt{n}\,|T_n - \theta| \leqq t\} = 1 - 2\sum_{m=1}^{\infty} \frac{1}{4m^2 - 1} e^{-8m^2t^2}. \tag{4.8}
$$

Of course, in this simple parametric example there are estimators of order $1/n$ in probability; in *estimation* problems, the minimum distance method is most useful in examples of a more nonparametric nature, where it often yields consistent estimators when other methods do not.

4.2. It is interesting to note that in the example of Section 4.1 and other similar cases it is simple to design along the lines of Section 2 a test of whether or not the unknown d.f. belongs to the specified class, where the limiting distribution of the test criterion when the null hypothesis is true is already known. These are the so-called "irregular" cases of estimation where an estimator of the unknown

parameter(s) indexing the class exists whose deviation from the true parameter value is of lower order in probability than the usual $1/\sqrt{n}$ encountered in "regular" cases; for example, of order $1/n$ for the rectangular distribution with unknown location and/or range, or for the exponential distribution with known scale but unknown location parameter. For the sake of definiteness, we fix our attention on a Kolmogoroff-type criterion for the latter example, although the result applies equally to other distributions and other criteria (ω^2-type tests with different "weight"-functions, etc). Thus, the problem is to test, on the basis of n observations, whether or not the true d.f. is of the form

$$0, \quad x < \theta, \qquad 1 - e^{-(x-\theta)}, \quad x \geqq \theta,$$

for some real θ. Let $T_n = \min(x_1, \cdots, x_n)$. Then $P_\theta\{\lim_{n\to\infty}\sqrt{n}(T_n - \theta) = 0\} = 1$. Hence, if we compare the sample d.f. with the exponential c.d.f. as estimated by using T_n, by computing

$$B_n = \sqrt{n} \sup_{0<r<1} \left| F_n\left(T_n + \log \frac{1}{1-r}\right) - r \right|,$$

we may conclude that, when the null hypothesis is true, B_n has the same limiting distribution as the Kolmogoroff statistic.

Remarks on power like those of Section 5 apply also to the present case. The present remarks may also be modified to apply to situations where one but not all parameters have irregular estimators, for example, for the case of the exponential distribution with unknown location *and* scale. For the case of the rectangular distribution with unknown range studied in Section 4.1, it seems intuitively reasonable that a test constructed in the manner of the present section may be more powerful than the one considered there.

5. Asymptotic power of the tests of normality. The results of this section are carried out for the tests of normality mentioned in Section 2, but the remarks below concerning v_n may also be carried through for the minimum distance test, the test of Section 3, and in many other examples, and the remarks concerning v_n and w_n may be extended to many other "distance" criteria.

First we consider the test of size α based on v_n. The critical region is of the form $\{v_n > b(\alpha)/\sqrt{n}\}$, where $b(\alpha)$ is a constant, except for terms of lower order in n. Suppose $G(x) \equiv R_0(x)$, and that $\delta(R_0, N^{**}) = d/\sqrt{n}$. From the theorem of Kolmogoroff [4] we have that $\delta(R_0, G_n^*)$ is of the order $1/\sqrt{n}$ in probability (uniformly in R_0). Hence, for $0 < \beta < 1$ there is a number $d^* = d^*(\alpha, \beta)$ such that $d > d^*$ implies that $\delta(G_n^*, N^{**}) > b(\alpha)/\sqrt{n}$ with probability $\geqq 1 - \beta$. From the definition of $\delta(G_n^*, N^{**})$ we have

$$\delta(G_n^*, N(y \mid \bar{x}, s^2)) \geqq \delta(G_n^*, N^{**}).$$

Thus, if we are using the test of size α, the power is at least $1 - \beta$ for any alternative R_0 whose distance from N^{**} is $\geqq d^*(\alpha, \beta)/\sqrt{n}$, and the power of the test at R_0 approaches one as $\sqrt{n}\delta(R_0, N^{**})$ increases indefinitely. This is a re-

sult of the same order as obtains for testing goodness of fit of a simple hypothesis by use of Kolmogoroff's distribution (Section 1).

We now consider the "ω^2-type" test criterion w_n.[5] We consider the function

$$\gamma(F, H) = \left\{\int_{-\infty}^{\infty} [F(x) - H(x)]^2 \, d\left(\frac{F(x) + H(x)}{2}\right)\right\}^{1/2} \tag{5.1}$$

as a possible measure of discrepancy between two d.f.'s. This measure has been used by Lehmann [14] and others. (We remark that γ is not a metric, since it does not satisfy the triangle inequality. Another undesirable property of γ is that the discrepancy between the d.f.'s of two random variables X and Y may not be the same as that between the d.f.'s of $-X$ and $-Y$. Also, the failure of the formula for integration by parts in expressions like (5.1) necessitates slight complications, for example, in the second following paragraph. Neither of the last two difficulties is present if both d.f.'s are continuous. If the d.f.'s have jumps which are nowhere dense, one could eliminate these last difficulties by redefining γ, for example, by replacing each jump by a constant density over an interval of width ϵ about the jump and letting $\epsilon \to 0$ after integrating. The development which follows would not be materially altered by such a change in the definition of γ.)

If F and H are continuous,

$$\int_{-\infty}^{\infty} [F(x) - H(x)]^2 \, d[F(x) - H(x)] = 0,$$

and the integration in (5.1) may be carried out with respect to either F or H instead of $\frac{1}{2}(F + H)$. Hence, if $G(x) \equiv R_0(x)$ is continuous,

$$\begin{aligned} w_n^{1/2} &= \left\{\int [G_n^*(x) - N(x \mid \bar{x}, s^2)]^2 \, d_x N(x \mid \bar{x}, s^2)\right\}^{1/2} \\ &\geqq \left\{\int [R_0(x) - N(x \mid \bar{x}, s^2)]^2 \, d_x N(x \mid \bar{x}, s^2)\right\}^{1/2} \\ &\quad - \left\{\int [R_0(x) - G_n^*(x)]^2 \, d_x N(x \mid \bar{x}, s^2)\right\}^{1/2} \\ &\geqq \gamma(R_0, N^{**}) - \delta(R_0, G_n^*), \end{aligned} \tag{5.2}$$

where

$$\gamma(R_0, N^{**}) = \inf_{N \in N^{**}} \gamma(R_0, N^{**}).$$

Now, $\delta(R_0, G_n^*)$ is of order $1/\sqrt{n}$ in probability (uniformly in R_0) and the critical region based on w_n is of the form $\{\sqrt{w_n} > c(\alpha)/\sqrt{n}\}$, where $c(\alpha)$ is constant except for terms of lower order in n. Hence, using (5.2), an argument like that

[5] In an unpublished manuscript, T. W. Anderson considers similar criteria for testing a simple hypothesis, and obtains similar results on asymptotic power. See also his abstract (*Ann. Math. Stat.*, Vol. 25 (1954), p. 174).

of the previous paragraph shows that there is a value $d'(\alpha, \beta)$ such that the test of size α has power $\geqq 1 - \beta$ for any continuous alternative R_0 for which $\gamma(R_0, N^{**}) \geqq d'(\alpha, \beta)/\sqrt{n}$.

We now consider what happens if R_0 is not continuous. It is easy to show, by consideration of the contribution to the integral of (5.1) at discontinuities of H, that, if F is continuous,

$$\int (F - H)^2 \, dF \geqq \tfrac{1}{4} \int (F - H)^2 \, dH,$$

so that

$$\int (F - H)^2 \, dF \geqq \tfrac{2}{5} \int (F - H)^2 \, d\left(\frac{F + H}{2}\right).$$

Hence, if R_0 is not continuous, the argument of the previous paragraph need only be altered by inserting the factor $\sqrt{2/5}$ before γ in the last expression of (5.2). The ensuing discussion of power then proceeds as before.

It is interesting to compare the measurements of distance δ and γ. Clearly, $\gamma(F, G) \leqq \delta(F, G)$ for all F,G. On the other hand, if there is a value x_0 for which $F(x_0) - G(x_0) = \delta$, it is clear from the monotonicity of F and G, using (5.1), that $[\gamma(F, G)]^2 \geqq \frac{1}{6}[\delta(F, G)]^3$. Thus, we have

$$\sqrt{1/6}\, \delta^{3/2} \leqq \gamma \leqq \delta, \tag{5.3}$$

where both equalities are attainable. We conclude that, whereas for any given $\beta > 0$, the power exceeds $1 - \beta$ for alternatives whose distance from N^{**} is of order $1/\sqrt{n}$ either according to δ for the test based on v_n or according to γ for the test based on w_n, the distance in terms of δ must be of order $1/\sqrt[3]{n}$ (in the worst case) to insure this for the latter test. For the former test, γ-distance of order $1/\sqrt{n}$ suffices.

We next verify the property of the χ^2-test relative to γ which was stated at the end of the third paragraph of the introduction. For brevity we shall use the notation of [3] without redefining symbols here; the reader may also refer to [3] for details of the argument which we omit. Suppose then that we are testing the hypothesis $G(x) = R(x) \equiv x$ for $0 \leqq x \leqq 1$ by means of the χ^2-test based on N observations and k_N intervals of equal length on the unit interval. If now

$$G(x) = F_{k_N}(x) \equiv \sum_{i=1}^{k_N} \frac{1}{k_N} \psi_x\left(\frac{2i - 1}{2k_N}\right)$$

(see Section 1 for the definition of ψ_x), then $\gamma(F_{k_N}, R) = 1/k_N\sqrt{6}$. Since F_{k_N} assigns the same probability as R to each of the k_N intervals, the power of the test against the alternative F_{k_N} is just the size of the test (assumed to be $< \frac{1}{2}$). We conclude that if the test gives power $\geqq \frac{1}{2}$ for all alternatives R^* satisfying $\gamma(R^*, R) \geqq \Gamma_N$, then we must have

$$k_N > 1/\Gamma_N\sqrt{6}. \tag{5.4}$$

(This is not the best possible inequality, but it suffices for our proof.) Consider now the distribution function

$$(5.5) \qquad H_a(x) = \begin{cases} (1 + 2a)x, & 0 \leqq x \leqq \frac{1}{2}, \\ 2a + (1 - 2a)x, & \frac{1}{2} \leqq x \leqq 1, \end{cases}$$

where $0 < a < \frac{1}{2}$.

A simple computation shows that $\gamma(H_a, R) = a/\sqrt{3} = \gamma_a$, say, and that (assuming for simplicity that k_N is even) when $G = H_a$ we have $\sum p_i^2 = (1 + 4a^2)/k_N = (1 + 12\gamma_a^2)/k_N$. Hence, the function ψ of [3] is given, when $G = H_a$, by

$$(5.6) \qquad \sigma'\psi(k_N) = 12(N - 1)\gamma_a^2 - C\sqrt{2(k_N - 1)}.$$

In order that the χ^2-test based on k_N intervals have power $\geqq \frac{1}{2}$ for all alternatives R^* with $\gamma(R^*, R) \geqq \Gamma_N$, it is necessary that the expression (5.6) be $\geqq 0$ asymptotically when we put $\gamma_a = \Gamma_N$, and that (5.4) be satisfied. We thus obtain, when N is large,

$$(5.7) \qquad \Gamma_N \geqq C'N^{-2/5}$$

where C' is a positive constant. From the result of [3] and the fact that $\gamma \leqq \delta$, we see that the reverse inequality to (5.7) is (for a different C') also true. Thus, $N^{-2/5}$ is indeed the smallest order of Γ_N which will give appreciable power for all alternatives R^* with $\delta(R, R^*) \geqq \Gamma_N$.

We shall now summarize the results proved thus far in this section. It is not known how the power function of the χ^2-test for composite hypotheses behaves, but it is plausible that the power function when testing a composite hypothesis by means of the χ^2-test (in any of its variations) is no better (in the sense we have used in measuring the goodness of a power function) than when testing a simple hypothesis. We have shown that *if* Γ *and* Δ *are small and the* χ^2*-test of size* $< \frac{1}{2}$ *of a simple hypothesis* $G = R$ *requires* N *observations to insure power* $\geqq \frac{1}{2}$ *at alternatives* R^* *for which* $\gamma(R^*, R) = \Gamma$ *(or* $\delta(R^*, R) = \Delta$*), so that* $N = C_1\Gamma^{-5/2}$ *(or* $N = C_2\Delta^{-5/2}$*), then the numbers of observations required by the Kolmogoroff and* ω^2 *tests to achieve the same minimum power at* Γ *(or* Δ*) are at most*

$$\textit{Kolmogoroff}: \quad n = C_3N^{4/5} = C_4\Gamma^{-2} \qquad (n = C_5N^{4/5} = C_6\Delta^{-2})$$

$$\omega^2: \quad n = C_7N^{4/5} = C_8\Gamma^{-2} \qquad (n = C_9N^{6/5} = C_{10}\Delta^{-3}).$$

The numbers of observations n required by the tests based on v_n *and* w_n *in testing composite hypotheses about parametric families are the same functions of* Γ *or* Δ *as for the Kolmogoroff and* ω^2 *tests of simple hypotheses.* The test based on v_n may thus be expected to be superior to the χ^2-test in the sense of both γ and δ, and that based on w_n may be expected to be superior in the sense of γ, at least for large N.

It might be supposed that the χ^2-test would show up to better advantage relative to the test based on v_n in terms of a metric like

$$\eta(R_0, N^{**}) = \inf_{N \epsilon N^{**}} \int d\, |R_0 - N|.$$

However, this is not so; for fixed n, even if $\eta(R_0, N^{**})$ is near its maximum of 2, neither the best χ^2-test of [3] nor any of the other tests we have mentioned need have appreciable minimum power ($\geq \frac{1}{2}$), and $\delta(R_0, N^{**})$ and $\gamma(R_0, N^{**})$ can be arbitrarily small. In fact, it is easy to see that no test can have the infimum of its power function over all alternatives R_0 with $\eta(R_0, N^{**}) = C > 0$ greater than the size of the test. (If N^{**} were a simple hypothesis, this would still be true.) In order better to compare the behavior of tests in terms of the metric η, we might therefore restrict our consideration to alternatives R_0 belonging to some regular class, for example, the class of d.f.'s with densities which cross that of each member of N^{**} at most M times. Under such a comparison the test based on v_n may be shown to be superior to that based on the χ^2-test, in the same sense as under our previous comparison.

The discussion of this section suggests very strongly that there is a "natural" distance with respect to which the power characteristics of a particular "distance" test criterion should be measured, and that a comparison of the power of such tests in terms of their own and other distances indicates that tests corresponding to strong metrics have the best global power characteristics. It is hoped to investigate this idea further.

6. Numerical results. In this section we list some pertinent experimental and computational results. The main purpose of the sampling experiments was to obtain estimates of the distributions of nw_n and $\sqrt{n}v_n$ which may be used in applications to test normality. As a check on the sampling experiments the same data were used to compute experimentally the d.f. of $\bar{x}$ and the d.f. of $\sqrt{n}D_n$ with n large, D_n being defined by

$$D_n = \delta(L(x), L_n^*(x))$$

where $L_n^*(x)$ is the empiric d.f. of n independent chance variables with the d.f. $L(x)$. Also, as another check on the experimentally obtained d.f. of nw_n, an approximation to the d.f. of W was computed, using (2.9).

Define

$$\omega_n^2 = \int (L(x) - L_n^*(x))^2\, dL(x).$$

The sampling experiments were conducted using 400 samples of size $n = 100$ and 400 samples of size $n = 25$, of random standard (mean zero, variance one) normal deviates from the well-known Rand Corporation series. From each sample the values of the sample mean ($\bar{x}$), sample variance (s^2), v_n, w_n, D_n, and ω_n^2, were computed. Thus, there were obtained 400 "observations" on $\sqrt{n}v_n$ and nw_n for $n = 25$ and $n = 100$, and the sample d.f.'s based on these observations serve as estimates of the d.f.'s of $\sqrt{n}v_n$ and nw_n. The known d.f. of $\bar{x}$

and the known d.f. of $\sqrt{n}D_n$ were compared with the experimentally obtained d.f.'s of $\bar{x}$ and $\sqrt{n}D_n$ as a check on the experimentally obtained d.f.'s of $\sqrt{n}v_n$ and nw_n.

It was found that the experimentally obtained d.f. of $\bar{x}$ agreed well with the known d.f., and the experimentally obtained d.f. of $\sqrt{n}D_n$ agreed well with the d.f. of $\sqrt{n}D_n$ as tabulated in [15]. In each case, the maximum difference between the experimentally obtained and the known actual d.f. was found to be small according to the tables [15]; this is the basis for saying the agreement was good. The agreement between the limiting d.f. of $n\omega_n^2$ (tabulated in [16]) and the experimentally obtained d.f. was found to be fairly close for $n = 100$ but not close for $n = 25$; the upper tails of the distributions seem to be the parts which come into agreement most rapidly with n. The distributions of $\sqrt{n}v_n$ and nw_n are concentrated closer to the origin than those of $\sqrt{n}D_n$ and $n\omega_n^2$, respectively. This is not entirely surprising since the covariance function (2.1) is smaller than that for the process corresponding to D_n and ω_n^2 (namely, min $(s, t) - st$). Scatter diagrams of the sample values of $(\sqrt{n}D_n, n\omega_n^2)$ indicate that the regression of $n\omega_n^2$ on $\sqrt{n}D_n$ is roughly parabolic with the conditional variance of $n\omega_n^2$ increasing with the value of $\sqrt{n}D_n$; this is not entirely surprising in view of the way in which ω_n^2 and D_n are computed from a sample. Although v_n and w_n are correlated with D_n and ω_n^2, they seem to be more strongly related to $\bar{x}$.

In Tables I and II, respectively, are given the estimates of the d.f.'s of $\sqrt{n}v_n$ and nw_n; commonly used percentage points are listed for convenience in Table III.

We now turn to the limiting distribution of nw_n. The first eight zeros of $D(\mu)$ are alternately zeros of $D_1(\mu)$ and $D_2(\mu)$. Their values, obtained by numerical computation of these functions for various values of μ, are 7.38, 8.62, 13.66, 15.14, 19.91, 21.52, 26.16, 27.87. The corresponding values of λ_j for $1 \leqq j \leqq 8$ are .01836, .01346, .00536, 00436, .00252, .00216, .00146, .00129. It is to be noted that λ_{2j-1} and λ_{2j} are both approximately c/j^2 (the corresponding property for the eigenvalues arising in the computation of $n\omega_n^2$ is that $\lambda_j = 1/\pi^2 j^2$). This (inferred) speed of convergence of the λ_j to 0 implies that the distribution of $W^* = \sum_1^4 \lambda_j \gamma_j$ should be a fairly good approximation to that of $W = \sum_1^\infty \lambda_j \gamma_j$,

TABLE I

Estimate of $Q_n(x) = P\{\sqrt{n}v_n \leq x\}$.

x	$Q_{25}(x)$	$Q_{100}(x)$	x	$Q_{25}(x)$	$Q_{100}(x)$
.30	0	0	.75	.8100	.8425
.35	.0125	.0025	.80	.8600	.9025
.40	.0550	.0250	.85	.9125	.9350
.45	.1500	.0800	.90	.9500	.9600
.50	.2525	.1975	.95	.9725	.9775
.55	.3675	.3300	1.00	.9850	.9900
.60	.5025	.4775	1.05	.9950	.9925
.65	.6400	.6750	1.10	1.0000	.9975
.70	.7225	.7325	1.15	1.0000	1.0000

TABLE II

Estimate of $R_n(x) = P\{nw_n \leqq x\}$ and distribution function $H(x) = P\{W^* \leqq x\}$.

x	$R_{25}(x)$	$R_{100}(x)$	$H(x)$
.01	0	0	.108
.02	.0375	.0225	.297
.03	.1325	.1400	.471
.04	.2125	.3175	.609
.05	.4300	.4775	.712
.06	.5825	.6250	.787
.07	.6625	.7175	.843
.08	.7250	.7850	.883
.09	.7925	.8600	.914
.10	.8400	.8900	.937
.11	.8975	.9175	.953
.12	.9150	.9350	.966
.13	.9425	.9700	.975
.14	.9525	.9825	.982
.15	.9650	.9850	.987
.16	.9700	.9850	.997
.17	.9800	.9875	
.18	.9875	.9875	
.19	.9875	.9900	
.20	.9900	.9950	
.21	.9950	.9950	
.22	.9975	.9975	
.23	1.0000	1.0000	

TABLE III

Estimates of common percentage points of Q_n and R_n.

p	$Q_{25}^{-1}(p)$	$Q_{100}^{-1}(p)$	$R_{25}^{-1}(p)$	$R_{100}^{-1}(p)$
.20	.7435	.729	.0909	.0824
.10	.8225	.797	.1145	.1019
.05	.8980	.878	.1352	.1240
.02	.9685	.954	.1671	.1386
.01	1.0145	.989	.1957	.1859
.005	1.0465	1.062	.2053	.1957

and the distribution of W^* was therefore computed, using the method of [17] (here $\lambda_j\gamma_j$ is the R_j of (2.9)). The results are given in the last column of Table II.

From the fact noted above regarding the speed of convergence of the d.f. of $n\omega_n^2$, and the fact that Table II indicates (in the difference $R_{100}(x) - R_{25}(x)$) a much slower approach to the limiting distribution for $R_n(x)$ than that noted in our experiment for the ω_n^2-distribution, we would expect that the d.f. $H(x)$ of W^* should lie above the estimates $R_n(x)$ of Table II, being close to $R_{100}(x)$ only

in the upper tail. This is what the last column of Table II actually shows, and an idea of how good the agreement is in the tail may be obtained by computing $M(x) = 20[H(x)]^{-1/2}[H(x) - R_{100}(x)]$; for $x = .13, .14, .15, .20$, one obtains $M(x) < 1$, which indicates very good agreement. Thus, in applications where n is large, it seems reasonable to use the last column of Table II, especially in the upper tail (which is the region that matters for statistical tests).

We are indebted to Prof. R. J. Walker and to Mr. R. C. Lesser of the Cornell Computing Center for carrying out the sampling and the computation of the λ_j's. The experimental data are available in the files of the Cornell Computing Center.

REFERENCES

[1] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[2] H. Chernoff and E. L. Lehmann, "The use of maximum likelihood estimates in χ^2 tests for goodness of fit," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 579–586.

[3] H. B. Mann and A. Wald, "On the choice of the number of class intervals in the application of the chi square test," *Ann. Math. Stat.*, Vol. 13 (1942), pp. 306–317.

[4] A. N. Kolmogoroff, "Sulla determinazione empirica di una leggi di distribuzione," *Giorn. Ist. Ital. Attuari*, Vol. 4 (1933), pp. 83–91.

[5] J. Wolfowitz, "Consistent estimators of the parameters of a linear structural relation," *Skand. Aktuarietids.*, Vol. 35 (1952), pp. 132–151.

[6] J. Wolfowitz, "Estimation by the minimum distance method," *Ann. Inst. Stat. Math., Tokyo*, Vol. 5 (1953), pp. 9–23.

[7] J. Wolfowitz, "Estimation by the minimum distance method in nonparametric stochastic difference equations," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 203–217.

[8] J. Wolfowitz,[6] "Estimation of the components of stochastic structures," *Proc. Nat. Acad. Sci., U.S.A.*, Vol. 40 (1954), pp. 602–606.

[9] H. B. Mann and A. Wald, "On stochastic limit and order relationships," *Ann. Math. Stat.*, Vol. 14 (1943), pp. 217–226.

[10] M. Kac, "On some connections between probability theory and differential and integral equations," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 180–215.

[11] M. Kac and A. J. F. Siegert, "On the theory of noise in radio receivers with square law detectors," *J. Applied Physics*, Vol. 18 (1947), pp. 383–397.

[12] M. Kac and A. J. F. Siegert, "An explicit representation of stationary Gaussian processes," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 438–442.

[13] J. L. Doob, "Heuristic approach to the Kolmogorov-Smirnoff theorems," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 393–403.

[14] E. L. Lehmann, "Consistency and unbiasedness of certain nonparametric tests," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 165–179.

[15] Z. W. Birnbaum, "Numerical tabulation of the distribution of Kolmogoroff's statistic for finite sample size," *J. Amer. Stat. Assn.*, Vol. 47 (1952), pp. 425–441.

[16] T. W. Anderson and D. A. Darling, "Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 193–212.

[17] H. Robbins and E. J. G. Pitman, "Application of the method of mixtures of quadratic forms in normal variates," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 552–560.

[18] D. A. Darling, "The Cramér-Smirnov test in the parametric case" (abstract), *Ann. Math. Stat.*, Vol. 24 (1953), p. 493.

[6] The sentence at the bottom of page 602 of [8] which reads, "It is well known that $F(z)$ determines α . . ." should read, "It is well known that the distribution of (x_i, x_{i+1}) determines α. . . ."

SOME CLASSES OF PARTIALLY BALANCED DESIGNS[1]

By R. C. Bose and W. H. Clatworthy[2]

University of North Carolina

1. Summary. Incomplete block designs with a few replications are of practical interest to experimenters. Partially balanced incomplete block (PBIB) designs with two associate classes, and $k > r = 2$, were studied by one of the authors [1]. The present paper extends this investigation to the case $k > r$ with $\lambda_1 = 1$ and $\lambda_2 = 0$. It is shown that the parameters of all PBIB designs in this case are given by (4.27) and thus depend upon three integral parameters k, r, and t, with the additional restrictions that

(i) $1 \leqq t \leqq r$,

(ii) $rk(r - 1)(k - 1)/t(k + r - t - 1)$ is a positive integer.

For the particular case $r = 3$ it is shown that all designs with $t = 2$ or 3 necessarily exist, but if $t = 1$, then the only possible value of $k > r$ is 5. However designs with parameters (4.27) with $r = 3$, $t = 1$, and $k = 2$ or 3 are also combinatorially possible though they do not belong to the class $k > r$.

Interesting by-products of this study are a lemma and five corollaries which give an insight into the structure of PBIB designs with $\lambda_1 = 1$ and $\lambda_2 = 0$, and $p_{11}^1 = k - 2$, no special assumptions being made regarding r and k.

2. Introduction. PBIB designs with m associate classes ($m \geqq 1$) were introduced by Bose and Nair [2]. Balanced incomplete block designs and square lattices were included as special cases. Nair and Rao [5] broadened the definition so as to further include cubic and other higher dimensional lattices. Bose and Shimamoto [3] have rephrased the definition so as to stress the fact that the relations between the treatments are determined only by the parameters n_i and p_{jk}^i with $i, j, k = 1, 2, \cdots, m$. For the special case of two associate classes ($m = 2$) the Bose and Shimamoto definition is substantially as follows.

A PBIB block design with two associate classes is an arrangement of v treatments (or varieties) in b blocks such that:

(i) Each of the v treatments is replicated r times in b blocks each of size k, and no treatment appears more than once in any block.

(ii) There exists a relationship of association between every pair of the v treatments satisfying the following conditions:

(a) Any two treatments are either first or second associates.

(b) Each treatment has n_1 first and n_2 second associates.

(c) Given any two treatments which are ith associates, the number p_{jk}^i of treatments common to the jth associates of the first and the kth associates

---

Received July 28, 1954.

[1] This work was supported in part by the United States Air Force under Contract No. AF 18(600)-83 monitored by the Office of Scientific Research.

[2] Now with the National Bureau of Standards, Washington, D. C.

of the second is independent of the pair of treatments with which we start. Furthermore, $p^i_{jk} = p^i_{kj}$ for $i, j, k = 1, 2$.

(iii) Any pair of treatments which are ith associates occur together in exactly λ_i blocks for $i = 1, 2$.

If either n_1 or n_2 assumes the value zero, one associate class does not exist. Hence, we shall require that both n_1 and n_2 be positive integers.

$$vr = bk, \tag{2.1}$$

$$v = n_1 + n_2 + 1, \tag{2.2}$$

$$\lambda_1 n_1 + \lambda_2 n_2 = r(k - 1), \tag{2.3}$$

$$p^1_{11} + p^1_{12} + 1 = p^2_{11} + p^2_{12} = n_1, \tag{2.4}$$

$$p^1_{21} + p^1_{22} = p^2_{21} + p^2_{22} + 1 = n_2, \tag{2.5}$$

$$n_1 p^1_{12} = n_2 p^2_{11}, \qquad n_1 p^1_{22} = n_2 p^2_{12}. \tag{2.6}$$

Furthermore, it was proved that if values are assigned to the parameters of the first kind (v, b, r, k, λ_1, λ_2, n_1, and n_2) satisfying (2.1), (2.2), and (2.3), then there is one independent parameter of the second kind (p^i_{jk} for $i, j, k = 1, 2$).

The parameters of the second kind will be exhibited as elements of two symmetric matrices

$$P_1 = \begin{pmatrix} p^1_{11} & p^1_{12} \\ p^1_{21} & p^1_{22} \end{pmatrix}, \qquad P_2 = \begin{pmatrix} p^2_{11} & p^2_{12} \\ p^2_{21} & p^2_{22} \end{pmatrix}. \tag{2.7}$$

Nair [6] established a necessary condition for PBIB designs having $k > r$. This condition was used by Bose [1] in exhausting the subclass of PBIB designs with two associate classes and $\lambda_1 = 1$ and $\lambda_2 = 0$, with $r = 2$. For the special case of PBIB designs with two associate classes with $\lambda_1 = 1$ and $\lambda_2 = 0$, and $k > r$, Nair's condition simplifies to

$$r p^1_{12} - (r - 1) p^2_{12} = r(r - 1), \tag{2.8}$$

a useful tool in the present investigation.

3. A less demanding definition for PBIB designs with two associate classes. We shall now show that for PBIB designs with two associate classes the Bose and Shimamoto definition is more demanding than it need be. To this end we establish two theorems.

THEOREM 3.1. *Let there exist a relationship of association between every pair among the v treatments satisfying the conditions:*

(a) *Any two treatments are either first or second associates.*

(b) *Each treatment has n_1 first and n_2 second associates.*

(c) *For any pair of treatments which are first associates, the number p^1_{11} of treatments common to the first associates of the first and the first associates of the second is independent of the pair of treatments with which we start.*

Then, for every pair of first associates among the v treatments the numbers p^1_{12}, p^1_{21}, *and* p^1_{22} *are constants, and* $p^1_{12} = p^1_{21}$.

PROOF. Let θ and ϕ be an arbitrarily chosen pair of first associates from among the v treatments. Let $p^1_{jk}(\theta, \phi)$ denote the number of treatments common to the jth associates of θ and to the kth associates of ϕ, for $j, k = 1, 2$. The n_1 first associates of θ are made up of ϕ, the $p^1_{11}(\theta, \phi)$ treatments which are first associates of θ as well as ϕ, and the $p^1_{12}(\theta, \phi)$ treatments which are first associates of θ but second associates of ϕ. Hence

$$1 + p^1_{11}(\theta, \phi) + p^1_{12}(\theta, \phi) = n_1 . \tag{3.1}$$

Likewise, classifying the first associates of ϕ, we have

$$1 + p^1_{11}(\theta, \phi) + p^1_{21}(\theta, \phi) = n_1 . \tag{3.2}$$

Similarly, the n_2 second associates of θ are made up of the $p^1_{21}(\theta, \phi)$ treatments which are second associates of θ and first associates of ϕ, and the $p^1_{22}(\theta, \phi)$ treatments which are second associates of both θ and ϕ. Hence

$$p^1_{21}(\theta, \phi) + p^1_{22}(\theta, \phi) = n_2 . \tag{3.3}$$

But by hypothesis, $p^1_{11}(\theta, \phi)$ is independent of the pair of treatments θ and ϕ and is p^1_{11}. Hence, from (3.1), (3.2), and (3.3) we obtain

$$p^1_{12}(\theta, \phi) = p^1_{21}(\theta, \phi) = n_1 - p^1_{11} - 1, \tag{3.4}$$

$$p^1_{22}(\theta, \phi) = n_2 - n_1 + p^1_{11} + 1. \tag{3.5}$$

Since θ and ϕ are an arbitrarily chosen pair of first associates, the relations (3.4) and (3.5) amount to a proof of the theorem.

THEOREM 3.2. *Let there exist a relationship of association between every pair among the v treatments satisfying the conditions:*

(a) *Any two treatments are either first or second associates.*

(b) *Each treatment has* n_1 *first and* n_2 *second associates.*

(c) *For any pair of treatments which are second associates, the number* p^2_{11} *of treatments common to the first associates of the first and the first associates of the second is independent of the pair of treatments with which we start.*

Then, for every pair of second associates among the v treatments the numbers p^2_{12}, p^2_{21}, *and* p^2_{22} *are constants, and* $p^2_{12} = p^2_{21}$.

Proof is similar to that of Theorem 3.1. In fact, it can be shown that

$$p^2_{12}(\theta, \phi) = p^2_{21}(\theta, \phi) = n_1 - p^2_{11}, \tag{3.6}$$

$$p^2_{22}(\theta, \phi) = n_2 - n_1 + p^2_{11} - 1, \tag{3.7}$$

where θ and ϕ are an arbitrarily chosen pair of second associates.

The question naturally arises whether one of the preceding theorems implies the other. The answer is no. Consider the following design with $v = 7$ treat-

ments in $b = 14$ blocks, each of size $k = 2$, and with each treatment replicated $r = 4$ times:

$$(0, 1), (1, 2), (2, 3), (3, 4), (4, 5), (5, 6), (6, 0),$$
$$(0, 3), (1, 4), (2, 5), (3, 6), (4, 0), (5, 1), (6, 2).$$

Each treatment of this design has $n_1 = 4$ first associates and $n_2 = 2$ second associates with $\lambda_1 = 1$ and $\lambda_2 = 0$. For any pair of treatments θ and ϕ which are second associates, $p_{11}^2(\theta, \phi) = 3$. Hence, this design satisfies Theorem 3.2 with $p_{12}^2 = p_{21}^2 = 1$ and $p_{22}^2 = 0$. However, for any two treatments α and β which are first associates, $p_{11}^1(\alpha, \beta) = 1$ or 2, and Theorem 3.1 is not satisfied.

Insofar as PBIB designs with two associate classes are concerned, the consequence of Theorems 3.1 and 3.2 is that the definition of a PBIB design given by Bose and Shimamoto demands more than is needed. For PBIB designs with two associate classes, a less demanding definition could be formed by replacing condition (c) of (ii) in Section 2 by

(c′) For any pair of the v treatments which are ith associates, the number, p_{11}^i for $i = 1, 2$, of treatments common to the first associates of the first and the first associates of the second is independent of the pair of treatments with which we start.

This new definition and Theorems 3.1 and 3.2 are then equivalent to the Bose and Shimamoto definition. Under the Bose and Shimamoto definition, to prove that an arrangement of objects is a PBIB design, it is necessary, insofar as (c) of (ii) is concerned, to show the constancy of all eight parameters p_{jk}^i for $i, j, k = 1, 2$, and also to show that the equalities $p_{jk}^i = p_{kj}^i$ hold for $j \neq k$ and $i, k = 1, 2$. By use of Theorems 3.1 and 3.2, with regard to the parameters of the second kind, it is necessary only to show the constancy of p_{11}^1 and p_{11}^2.

4. Complete enumeration of PBIB designs with two associate classes and $k > r \geqq 2$, with $\lambda_1 = 1$ and $\lambda_2 = 0$. We first establish the following useful lemma.

LEMMA 4.1. *For any* PBIB *design with two associate classes and* $k > r \geqq 2$, with $\lambda_1 = 1$ *and* $\lambda_2 = 0$,

$$k - 2 \leqq p_{11}^1 \leqq (k - 2) + (r - 1)^2. \tag{4.1}$$

PROOF. The left portion of (4.1) is obvious; consider the right portion. Let the treatments θ and ϕ be first associates. Then they occur together in exactly one block which we shall denote by $B(\theta, \phi)$. There are $k - 2$ other treatments in this block which are first associates of θ as well as ϕ. Since $\lambda_1 = 1$, both θ and ϕ cannot occur together in any other block. Denote the $r - 1$ blocks in which θ but not ϕ occurs by $B_i(\theta)$, for $i = 1, 2, \cdots, r - 1$, and similarly the $r - 1$ blocks in which ϕ but not θ occurs by $B_j(\phi)$, for $j = 1, 2, \cdots, r - 1$. If a treatment does not occur in $B(\theta, \phi)$ but is a first associate of θ as well as ϕ, it must occur exactly once in the blocks $B_i(\theta)$, and exactly once in the blocks $B_j(\phi)$. But the block $B_j(\phi)$ cannot have more than one treatment in common with any

of the blocks $B_1(\theta), B_2(\theta), \cdots, B_{r-1}(\theta)$. Hence $B_j(\phi)$ cannot contain more than $r-1$ first associates of θ. This holds for $j = 1, 2, \cdots, r-1$. Therefore there cannot exist more than $(r-1)^2$ treatments which occur once among the blocks $B_i(\theta)$ and once among the blocks $B_j(\phi)$. Thus p_{11}^1 cannot exceed $(k-2) + (r-1)^2$, which proves the lemma.

We shall now obtain the combinatorial parameters of all designs belonging to the class under consideration. From (2.4) and (2.6) we obtain

$$n_1 p_{12}^1 + n_2 p_{12}^2 = n_1 n_2 . \tag{4.2}$$

Solving (2.8) and (4.2) simultaneously, we have

$$p_{12}^1 = n_2(r-1)(r+n_1) \,/\, [n_1(r-1) + n_2 r] \tag{4.3}$$

$$p_{12}^2 = r n_1(n_2 - r + 1) \,/\, [n_1(r-1) + n_2 r]. \tag{4.4}$$

From (2.2) and (2.3) we get

$$n_1 = r(k-1) \tag{4.5}$$

$$v = n_2 + 1 + r(k-1). \tag{4.6}$$

From (2.1) and (4.6) it follows that

$$b = r^2 + (r/k)(n_2 - r + 1). \tag{4.7}$$

Since both b and r^2 must be integral, we set

$$r(n_2 - r + 1) = sk \tag{4.8}$$

where s is an integer. Then, from (4.8), (4.7), (2.1), (4.3), (4.4), and (4.5) we get

$$n_2 = sk/r + r - 1, \tag{4.9}$$

$$b = r^2 + s, \tag{4.10}$$

$$v = k(r^2 + s)/r, \tag{4.11}$$

$$p_{12}^1 = (r-1)k - r(r-1)^2(k-1) \,/\, [s + r(r-1)], \tag{4.12}$$

$$p_{12}^2 = r(k-1) - r^2(r-1)(k-1) \,/\, [s + r(r-1)]. \tag{4.13}$$

From (4.12) and (4.13) it is seen that $r(r-1)^2(k-1)$ and $r^2(r-1)(k-1)$ must both be integral multiples of $s + r(r-1)$. Hence, their difference must also be divisible by $s + r(r-1)$. Therefore, we introduce an auxiliary integral parameter t defined by

$$t = r(r-1)(k-1) \,/\, [s + r(r-1)]. \tag{4.14}$$

Then

$$s = r(r-1)(k-t-1)/t, \qquad t \neq 0. \tag{4.15}$$

From (4.9) through (4.13) and (4.15) it follows that

$$v = k[(r-1)(k-1)+t]/t, \tag{4.16}$$

$$b = r[(r-1)(k-1)+t]/t, \tag{4.17}$$

$$n_2 = (r-1)(k-1)(k-t)/t, \tag{4.18}$$

$$p^1_{12} = (r-1)(k-t), \tag{4.19}$$

$$p^2_{12} = r(k-t-1). \tag{4.20}$$

From (4.19), (4.20), (2.4), and (2.5) we get

$$p^1_{11} = t(r-1)+k-r-1, \tag{4.21}$$

$$p^1_{22} = (r-1)(k-t)\,(k-t-1)/t, \tag{4.22}$$

$$p^2_{11} = rt, \tag{4.23}$$

$$p^2_{22} = [(r-1)(k-1)(k-2t)+t(rt-k)]/t. \tag{4.24}$$

Applying Lemma 4.1 to (4.21), we obtain

$$1 \leqq t \leqq r, \tag{4.25}$$

a most useful set of bounds on the integral parameter t. It is now seen that the divisor, $s + r(r-1)$, in (4.12) produces no difficulty because

$$(r-1)(k-1) \leqq s + r(r-1) \leqq r(r-1)(k-1), \qquad k > r \geqq 2. \tag{4.26}$$

In summary, all PBIB designs with two associate classes belonging to the class characterized by $k > r \geqq 2$, with $\lambda_1 = 1$ and $\lambda_2 = 0$, are obtainable from

$$\left\{\begin{aligned} &v = k[(r-1)(k-1)+t]/t, \qquad r = r, \qquad \lambda_1 = 1,\\ &b = r[(r-1)(k-1)+t]/t, \qquad k = k, \qquad \lambda_2 = 0,\\ &n_1 = r(k-1), \qquad n_2 = (r-1)(k-1)(k-t)/t,\\ &P_1 = \begin{pmatrix} (t-1)(r-1)+k-2 & (r-1)(k-t) \\ (r-1)(k-t) & (r-1)(k-t)(k-t-1)/t \end{pmatrix},\\ &P_2 = \begin{pmatrix} rt & r(k-t-1) \\ r(k-t-1) & [(r-1)(k-1)(k-2t)+t(rt-k)]/t \end{pmatrix}, \end{aligned}\right. \tag{4.27}$$

where $1 \leqq t \leqq r$.

Connor and Clatworthy [4] have shown (Theorem 5.1) that for a PBIB design with two associate classes it is necessary (but not sufficient) that the quantities

$$\alpha_1 = [(v-1)(-\gamma+\sqrt{\Delta}+1)-2n_1]\,/\,2\sqrt{\Delta} \tag{4.28}$$

$$\alpha_2 = [(v-1)(\gamma+\sqrt{\Delta}+1)-2n_2]\,/\,2\sqrt{\Delta} \tag{4.29}$$

be positive integers, where

$$\gamma = p^2_{12} - p^1_{12}, \qquad \beta = p^1_{12} + p^2_{12}, \qquad \Delta = \gamma^2 + 2\beta + 1. \tag{4.30}$$

Thus for any design of (4.27) it is necessary that

$$\alpha_1 = rk(r - 1)(k - 1) / t(k + r - t - 1). \tag{4.31}$$

In the case of designs of three replications, there are only three series, one corresponding to each of the three permissible values of t. Setting $t = 1, 2,$ and 3 in (4.27) we obtain

$$\text{(4.32)}\begin{cases} v = k(2k-1), \quad r = 3, \quad \lambda_1 = 1, \quad n_1 = 3(k-1), \\ b = 3(2k-1), \quad k = k, \quad \lambda_2 = 0, \quad n_2 = 2(k-1)^2, \\ P_1 = \begin{pmatrix} k-2 & 2(k-1) \\ 2(k-1) & 2(k-1)(k-2) \end{pmatrix}, \quad P_2 = \begin{pmatrix} 3 & 3(k-2) \\ 3(k-2) & 2k^2 - 7k + 7 \end{pmatrix}; \end{cases}$$

$$\text{(4.33)}\begin{cases} v = k^2, \quad r = 3, \quad \lambda_1 = 1, \quad n_1 = 3(k-1), \\ b = 3k, \quad k = k, \quad \lambda_2 = 0, \quad n_2 = (k-1)(k-2), \\ P_1 = \begin{pmatrix} k & 2(k-2) \\ 2(k-2) & (k-2)(k-3) \end{pmatrix}, \quad P_2 = \begin{pmatrix} 6 & 3(k-3) \\ 3(k-3) & k^2 - 6k + 10 \end{pmatrix}; \end{cases}$$

$$\text{(4.34)}\begin{cases} v = k(2k+1)/3, \quad r = 3, \quad \lambda_1 = 1, \quad n_1 = 3(k-1), \\ b = 2k+1, \quad k = k, \quad \lambda_2 = 0, \quad n_2 = 2(k-1)(k-3)/3, \\ P_1 = \begin{pmatrix} k+2 & 2(k-3) \\ 2(k-3) & 2(k-3)(k-4)/3 \end{pmatrix}, \\ P_2 = \begin{pmatrix} 9 & 3(k-4) \\ 3(k-4) & (2k^2 - 17k + 39)/3 \end{pmatrix}. \end{cases}$$

In (4.34), k must be of the form $3p$ or $3p + 1$, p being an integer.

The designs given by (4.33) are the well known lattice designs of three replications. For all integral values of $k \geqq 3$ solutions exist.

The results of Reiss [7], together with those of Shrikhande [8], are sufficient to guarantee the existence of solutions for the designs of (4.34) as duals of the corresponding balanced incomplete block designs given by

$$v^* = 2k + 1, \qquad b^* = k(2k + 1)/3, \qquad k^* = 3, \qquad r^* = k, \qquad \lambda^* = 1, \tag{4.35}$$

where k is of the form $3p$ or $3p + 1$, p being a positive integer.

It follows from (4.31), that for designs of (4.32), $12/(k + 1)$ must be integral. The only permissible values of k are therefore 2, 3, 5, and 11.

The designs corresponding to $k = 2$ and 3 are known, the design corresponding to $k = 5$ is new, and the case $k = 11$ is impossible. We shall defer the discussion of these designs to Section 7.

5. The block structure of PBIB designs with two associate classes having $\lambda_1 = 1$ and $\lambda_2 = 0$, with $p_{11}^1 = k - 2$. First, we shall introduce the notation and

terminology used throughout the remainder of this paper. We shall then establish a lemma and five corollaries which provide necessary conditions for the existence of the designs we consider.

Consider any PBIB design with two associate classes, and having $\lambda_1 = 1$ and $\lambda_2 = 0$, with $p_{11}^1 = k - 2$. Let the Greek letter θ represent any treatment of the design. The $n_1 = r(k - 1)$ first associates of θ will be denoted by r Latin letters each bearing subscripts $1, 2, \cdots, k - 1$. The $k - 1$ first associates of θ appearing together with θ in a block will be denoted by the same letter, and first associates of θ appearing in different blocks with θ will be denoted by different letters. The n_2 second associates of θ will be denoted by the integers $1, 2, \cdots, n_2$. The n_1 first associates of θ will sometimes be referred to as *lettered treatments* and the n_2 second associates of θ as *numbered treatments*. The r blocks containing treatment θ will be referred to as the *θ-blocks*.

Likewise, the blocks containing the lettered treatments x_j, where $x = a$ or b or c, etc., and $j = 1, 2, \cdots, k - 1$, but not θ will be referred to as the *x-blocks*, and the r blocks containing the numbered treatment i, with $1 \leqq i \leqq n_2$, will be called the *i-blocks*. Blocks containing only first associates of θ or only second associates of θ will be called *pure blocks* while blocks containing both first and second associates of θ will be called *mixed blocks*.

In the special case of designs having $\lambda_1 = 1$ and $\lambda_2 = 0$, with $p_{11}^1 = k - 2$, it will be shown that there are exactly $r(r - 1)(k - 1)$ mixed blocks each of which contains one lettered treatment and $k - 1$ numbered treatments. Those $(r - 1)(k - 1)$ mixed blocks containing first associates of θ denoted by the same letter (and subscripts $1, 2, \cdots, k - 1$) will be called a *group of blocks*. There are r such groups of blocks. The group of mixed blocks containing the lettered treatment x with any subscript will be referred to as the *x-group* of blocks ($x = a$, b, c, etc.) Within the x-group of blocks there are $r - 1$ blocks containing the lettered treatment x_j. These $r - 1$ blocks will be referred to as the *x_j-set* of blocks.

LEMMA 5.1. *For a* PBIB *design with two associate classes having* $\lambda_1 = 1$ *and* $\lambda_2 = 0$, *with* $p_{11}^1 = k - 2$, *any two treatments appearing in different blocks containing a common treatment must be second associates of each other.*

PROOF. There is no loss in generality if the treatment common to the two blocks is taken as treatment θ. Let x_i and y_j, with $x \neq y$ and $i, j = 1, 2, \cdots, k - 1$, be any two treatments appearing in different blocks containing θ. Since $\lambda_1 = 1$ and $\lambda_2 = 0$, two treatments appearing in the same block are first associates of each other while any pair of treatments which do not occur together in any block are second associates of each other. The $k - 2$ other treatments appearing in the same block with the pair (θ, x_i) are first associates of both θ and x_i. Since $p_{11}^1 = k - 2$, all other first associates of θ (including y_j) must be second associates of x_i. This proves the lemma.

COROLLARY 5.1. *Under the conditions of Lemma* 5.1, *each mixed block contains only one lettered treatment, and the* $r(r - 1)(k - 1)$ *mixed blocks can be divided into* r *distinct groups, each group containing* $k - 1$ *sets and each set containing* $r - 1$ *blocks.*

Proof. Let the r θ-blocks of the design be

$$\begin{array}{ccccc} \theta & a_1 & a_2 & \cdots & a_{k-1} \\ \theta & b_1 & b_2 & \cdots & b_{k-1} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \theta & l_1 & l_2 & \cdots & l_{k-1}. \end{array}$$

Since $\lambda_1 = 1$ and $\lambda_2 = 0$, no pair of treatments can occur together in more than one block. By Lemma 5.1 any pair of lettered treatments x_i and y_j, with $x \neq y$, must be second associates of each other. Hence, no two lettered treatments can appear together in a block free from θ. Therefore the design must contain $r(r-1) \cdot (k-1)$ mixed blocks, each of which contains one lettered and $k-1$ numbered treatments. Since each lettered treatment must appear in $r-1$ blocks free from θ, there are $(r-1)(k-1)$ mixed blocks containing treatments $x_1, x_2, \cdots, x_{k-1}$. These blocks constitute the x-group of blocks, where $x = a, b, \cdots$, or l, and there are r distinct groups of blocks. Within the x-group there are $r-1$ blocks containing the lettered treatment x_i, and these we have called the x_i-set of blocks for $i = 1, 2, \cdots, k-1$. Each group of blocks obviously contains $k-1$ distinct sets of blocks.

Corollary 5.2. *Under the conditions of Lemma 5.1, the numbered treatments appearing in a group of blocks must all be distinct. Hence* $n_2 \geqq (r-1)(k-1)^2$. *If* $n_2 = (r-1)(k-1)^2$, *each group must contain precisely one complete replication of the numbered treatments.*

Proof. Since $\lambda_1 = 1$ and $\lambda_2 = 0$, no numbered treatment can appear in two blocks belonging to the same set. Suppose the same numbered treatment j, for $j = 1, 2, \cdots, n_2$, appears in blocks belonging to different sets of the same group. Let the lettered treatment in one of the sets containing j be x_m and the lettered treatment in the other set containing j be x_n, with $m \neq n$ and $m, n = 1, 2, \cdots, k-1$. Now x_m and x_n are first associates of each other since they appear together in the same θ-block. But by Lemma 5.1, x_n and x_m must be second associates of each other since they appear in different blocks containing the common treatment j, a contradiction. Therefore no numbered treatment can appear in two or more blocks belonging to the same group. Hence the numbered treatments appearing in the same group of blocks must all be distinct. Since each mixed block contains $k-1$ numbered treatments, $(r-1)(k-1)^2$ distinct numbered treatments appear in a group of blocks. Since numbered treatments are second associates of θ, $n_2 \geqq (r-1)(k-1)^2$. If $n_2 = (r-1)(k-1)^2$, each group must obviously contain exactly one complete replication of the numbered treatments.

Corollary 5.3. *Under the conditions of Lemma 5.1, the mixed blocks contain* p_{11}^2 *complete replications of the numbered treatments. The pure blocks must contain* $r - p_{11}^2$ *complete replications of the numbered treatments.*

Proof. Consider treatment θ and an arbitrarily chosen numbered treatment i, for $i = 1, 2, \cdots, n_2$. Since θ and i are second associates of each other, and

since θ has only lettered treatments as first associates, treatment i must have exactly p_{11}^2 lettered first associates. Since lettered treatments appear only in θ-blocks and mixed blocks, and since no numbered treatments can appear in a θ-block, it follows by Corollary 5.1 that treatment i must appear in exactly p_{11}^2 mixed blocks. This is true for $i = 1, 2, \cdots, n_2$; therefore the mixed blocks contain exactly p_{11}^2 complete replications of the numbered treatments. Now, there are, in all, r complete replications of the numbered treatments. Therefore the pure blocks consist of $r - p_{11}^2$ complete replications of the numbered treatments.

Corollary 5.4. *Under the conditions of Lemma* 5.1, *two sets of blocks belonging to different groups must intersect in* $p_{11}^2 - 1$ *numbered treatments; that is, they must have* $p_{11}^2 - 1$ *numbered treatments in common.*

Proof. By Lemma 5.1 and Corollary 5.1, the pair of lettered treatments x_i and y_j, where $x \neq y$, appearing in two sets belonging to different groups are second associates of each other. Since x_i and y_j have θ common to their first associates but no common lettered first associates, they must have $p_{11}^2 - 1$ common numbered first associates. Since x_i and y_j appear only in the θ-blocks and in the sets under consideration, these two sets must contain the $p_{11}^2 - 1$ common numbered first associates of x_i and y_j. This proves the corollary.

Corollary 5.5. *Under the conditions of Lemma* 5.1 *a mixed block cannot intersect a set belonging to another group in more than one treatment. If* $n_2 = (r - 1) \cdot (k - 1)^2$, *then each mixed block must intersect each set of another group in exactly one numbered treatment.*

Proof. Suppose a mixed block intersects a set of another group in two or more treatments. Then consider a pair of numbered treatments common to the given mixed block and the intersected set of blocks belonging to another group. These two numbered treatments are first associates of each other since they occur together in a block. Their common first associates will include the $k - 2$ other treatments in the block containing them both and also the lettered treatment occurring in all blocks of the set intersected. Hence $p_{11}^1 \geqq k - 1$, in contradiction to the hypothesis. This proves the first part of the corollary.

Each group must contain a complete replication of numbered treatments when $n_2 = (r - 1)(k - 1)^2$, and we have proved that a mixed block cannot intersect a set of another group in more than one treatment. Therefore, when $n_2 = (r - 1)(k - 1)^2$, the $k - 1$ numbered treatments in a mixed block must be distributed one to each set in the other groups. This proves the second part of the corollary.

Definition. Two number-pairs may be defined to be *set compatible* if it is possible for them to occur in the same set without violating Lemma 5.1 and its corollaries.

6. The relationship of duality between members of a certain two-parameter series of PBIB designs. Let D be a design with a known solution. To form a design D^*, let the treatments of D be the blocks of D^*, and the blocks of D

be the treatments of D^*. The design D^* has been called the *dual* of design D. The process described is sometimes called *inverting* or *dualizing the* design D.

In general, the dual of a PBIB design is not itself a PBIB design. Inversion of a partially balanced design D can result in a dual design D^* having a different number of associate classes than design D. Inversion of a PBIB design D may yield a balanced incomplete block design.

Consider the two-parameter series of designs having

$$(6.1)\quad \begin{cases} v = k[(r-1)(k-1)+1], \quad r = r, \quad \lambda_1 = 1, \quad n_1 = r(k-1), \\ b = r[(r-1)(k-1)+1], \quad k = k, \quad \lambda_2 = 0, \quad n_2 = (r-1)(k-1)^2, \\ P_1 = \begin{pmatrix} k-2 & (r-1)(k-1) \\ (r-1)(k-1) & (r-1)(k-1)(k-2) \end{pmatrix}, \\ P_2 = \begin{pmatrix} r & r(k-2) \\ r(k-2) & k^2(r-1)-(3r-2)(k-1) \end{pmatrix}, \end{cases}$$

obtained by setting $t = 1$ in (4.27). Note that (6.1) satisfies the conditions of Lemma 5.1 and its five corollaries, with equality holding in Corollary 5.2.

Assuming the existence of a solution of some design belonging to (6.1), let us examine its dual. Let the parameters of the dual be distinguished by the asterisk as superscript. Clearly

$$(6.2)\quad \begin{aligned} v^* &= r[(r-1)(k-1)+1], \qquad r^* = k, \\ b^* &= k[(r-1)(k-1)+1], \qquad k^* = r. \end{aligned}$$

The r blocks of (6.1) containing treatment θ go over into r treatments appearing in a single block of the dual design. Since no pair of θ-blocks of (6.1) contain another common treatment, in the dual no treatment-pair corresponding to a pair of blocks containing θ can occur together in a block (excepting the one block corresponding to treatment θ). Hence

$$(6.3)\qquad \lambda_1^* = 1, \quad \lambda_2^* = 0.$$

Arbitrarily choose a block of the original design and a treatment within this block. There are $r-1$ other blocks containing this treatment. The arbitrarily chosen treatment may be called θ and the r blocks containing θ are then the θ-blocks. The arbitrarily chosen θ-block intersects each of the other $r-1$ θ-blocks in exactly one treatment. It also intersects each of the $(r-1)(k-1)$ blocks of one group in a single treatment, but it intersects no other blocks. Hence in the dual

$$(6.4)\qquad n_1^* = k(r-1).$$

These n_1^* blocks in the original design (treatments in the dual) we shall call the

first associates of the arbitrarily chosen block. Since the arbitrarily chosen block fails to intersect any of the blocks in the other $r - 1$ groups, in the dual

$$n_2^* = (k - 1)(r - 1)^2. \tag{6.5}$$

We shall call these n_2^* blocks the second associates of the arbitrarily chosen block.

Now consider any pair of blocks which are first associates. There is no loss in generality in assuming that they are the ith and jth blocks containing treatment θ for $i \neq j$; with $1 \leqq i$ and $j \leqq r$. The first associates of the ith block containing θ consist of all θ-blocks except the ith, and of all the blocks in the ith group (the ith group being the one whose lettered treatments occur in the ith θ-block). The first associates of the jth block containing θ consist of all θ-blocks except the jth, and of all blocks in the jth group. The blocks which are common to the first associates of the ith and jth θ-blocks are the other $r - 2$ θ-blocks. Hence,

$$p_{11}^{1*} = r - 2, \tag{6.6}$$

a constant for any particular design. Since the conditions of Theorem 3.1 are satisfied, it follows that p_{12}^{1*}, p_{21}^{1*}, and p_{22}^{1*} are constants, and from (3.4) and (3.5) we obtain

$$p_{12}^{1*} = p_{21}^{1*} = (r - 1)(k - 1), \tag{6.7}$$

$$p_{22}^{1*} = (r - 1)(r - 2)(k - 1). \tag{6.8}$$

Next consider a pair of blocks which are second associates of each other, that is, a pair of blocks containing no common treatment. There is no loss in generality if we take one of them to be the first θ-block and the other to be a block in the jth group, where $j \neq 1$. The first associates of the first θ-block are the $r - 1$ other θ-blocks and all the blocks in the first group. The first associates of a block in the jth group consist of the $r - 2$ other blocks in the same set of the jth group, one block (not the first) among the θ-blocks, and $k - 1$ blocks in each of the r groups excepting the jth. Hence

$$p_{11}^{2*} = k, \tag{6.9}$$

a constant for a particular design. Thus it is seen that the conditions of Theorem 3.2 are satisfied, and hence, p_{12}^{2*}, p_{21}^{2*}, and p_{22}^{2*} are constants. Furthermore, from (3.6) and (3.7) we obtain

$$p_{12}^{2*} = p_{21}^{2*} = k(r - 2), \tag{6.10}$$

$$p_{22}^{2*} = r^2(k - 1) - (3k - 2)(r - 1). \tag{6.11}$$

In the original design, any pair of blocks which are first associates interesect in a single treatment, so $\lambda_1^* = 1$, while any pair of blocks which are second associates fail to intersect at all, so $\lambda_2^* = 0$.

We have shown that the conditions (i), (ii), and (iii) stated in the definition of a PBIB design with two associate classes are satisfied. Hence, the dual of design (6.1) is a PBIB design with two associate classes having parameters

$$(6.12)\quad \begin{cases} v^* = r[(r-1)(k-1)+1], \quad r^* = k, \quad \lambda_1^* = 1, \quad n_1^* = k(r-1), \\ b^* = k[(r-1)(k-1)+1], \quad k^* = r, \quad \lambda_2^* = 0, \quad n_2^* = (k-1)(r-1)^2, \\ P_1^* = \begin{pmatrix} r-2 & (r-1)(k-1) \\ (r-1)(k-1) & (r-1)(r-2)(k-1) \end{pmatrix}, \\ P_2^* = \begin{pmatrix} k & k(r-2) \\ k(r-2) & r^2(k-1)-(3k-2)(r-1) \end{pmatrix}. \end{cases}$$

If in (6.12) we replace k with r^* and r with k^*, it is seen that the dual has the same form as the original design (6.1) and hence belongs to the same series. We have proved that the existence of the original design (6.1) implies the existence of (6.12) as the dual of (6.1). If we now assume the existence of (6.12), it can be shown by exactly the same type of argument that its dual exists and is given by (6.1). We have thus proved the following theorem.

THEOREM 6.1. *Between corresponding designs of* (6.1) *and* (6.12) *there exists a relationship of duality. They exist or fail to exist simultaneously.*

The theorem of duality implies that if a design belonging to one of the series (6.1) or (6.12) is impossible, then the corresponding design of the other series is also impossible. Suppose that there exists a solution of a design of one of the series, but that the corresponding design of the other series is impossible. Then by Theorem 6.1, upon inverting the design whose solution is known, we obtain the solution of the corresponding design of the other series. But this contradicts our assumption. Consequently, the first statement of this paragraph is true.

Actually, the series (6.1) and (6.12) are the same, since we may obtain (6.1) from (6.12) by interchanging r and k. However, the correspondence between designs of (6.1) and (6.12) gives a means of pairing off those designs which are duals of each other.

7. Constructions and impossibility proofs. Putting $k = 3$ in the two-parameter series of designs (6.1), we obtain the single-parameter series, with $r \geqq 2$,

$$(7.1)\quad \begin{cases} v = 3(2r-1), \quad r = r, \quad \lambda_1 = 1, \quad n_1 = 2r, \\ b = r(2r-1), \quad k = 3, \quad \lambda_2 = 0, \quad n_2 = 4(r-1), \\ P_1 = \begin{pmatrix} 1 & 2(r-1) \\ 2(r-1) & 2(r-1) \end{pmatrix}, \quad P_2 = \begin{pmatrix} r & r \\ r & 3r-5 \end{pmatrix}. \end{cases}$$

Putting $k = 3$ in the two-parameter series of designs (6.12), we obtain the single-parameter series

$$(7.2)\quad \begin{cases} v^* = r(2r-1), \quad r^* = 3, \quad \lambda_1^* = 1, \quad n_1^* = 3(r-1), \\ b^* = 3(2r-1), \quad k^* = r, \quad \lambda_2^* = 0, \quad n_2^* = 2(r-1)^2, \\ P_1^* = \begin{pmatrix} r-2 & 2(r-1) \\ 2(r-1) & 2(r-1)(r-2) \end{pmatrix}, \quad P_2^* = \begin{pmatrix} 3 & 3(r-2) \\ 3(r-2) & 2r^2-7r+7 \end{pmatrix}. \end{cases}$$

It follows from Section 6 that corresponding designs of (7.1) and (7.2) are duals of each other and exist or fail to exist simultaneously. Note that series (7.2) is identical with (4.32). It follows from (4.31) that designs of (7.2) corresponding to values of r other than 2, 3, 5, and 11 do not exist, and the same therefore must hold for designs of (7.1). In what follows we shall show that designs of (7.1) with $r \geqq 6$ are impossible, which rules out the case $r = 11$. We shall also give a construction for the case $r = 5$ for which the corresponding design for (7.2) can be obtained by dualization. The designs corresponding to $r = 2$ and 3 are known, and will be considered briefly at the end of this section.

We write the r θ-blocks of (7.1) as

$$\begin{array}{ccc} \theta & a_1 & a_2 \\ \theta & b_1 & b_2 \\ \cdots & \cdots & \cdots \\ \theta & l_1 & l_2 \end{array}$$

Each group consists of $2(r-1)$ blocks, the lettered treatments within a group being denoted by the same letter bearing the appropriate subscript. Each group contains two sets and each set contains $r-1$ blocks, the blocks within a set containing the same lettered treatment (Corollary 5.1). Since $n_2 = 4(r-1)$, each group must contain exactly one complete replication of the numbered treatments (Corollary 5.2). Without loss of generality, we may write the first group containing lettered treatments a_1 and a_2 as

$$\left.\begin{array}{ccc} a_1 & 1 & 2 \\ a_1 & 3 & 4 \\ a_1 & 5 & 6 \\ \cdots & \cdots & \cdots \\ a_1 & 2r-3 & 2r-2 \end{array}\right\} \quad r-1 \text{ blocks,}$$

$$\left.\begin{array}{ccc} a_2 & 2r-1 & 2r \\ a_2 & 2r+1 & 2r+2 \\ a_2 & 2r+3 & 2r+4 \\ \cdots & \cdots & \cdots \\ a_2 & 4r-5 & 4r-4 \end{array}\right\} \quad r-1 \text{ blocks.}$$

There is no loss in generality in writing the b_1-set as

$$\begin{array}{lcc}
b_1 & 1 & 2r-1 \\
b_1 & 3 & 2r+1 \\
b_1 & 5 & 2r+3 \\
\cdots & \cdots & \cdots \\
b_1 & 2r-3 & 4r-5
\end{array}$$

where the b_1-set contains only the odd-numbered second associates of θ (Corollaries 5.4 and 5.5). Then, the b_2-set must contain all the even-numbered second associates of θ (Corollary 5.2). The treatments $2, 4, 6, \cdots, 2r-2$ must occur one in each block of the b_2-set, the same being true for the $r-1$ treatments $2r, 2r+2, 2r+4, \cdots, 4r-4$.

We shall now show that the pairing off of the even-numbered treatments in the b_2-set is uniquely determined. Since no pair of numbered treatments appearing in the a_1- or a_2-set can occur together in any block of the b-, c-, $\cdots$, and l-groups ($\lambda_1 = 1$, $\lambda_2 = 0$, and Lemma 5.1), we form a lattice (Diagram 1) having horizontal coordinates $1, 2, 3, \cdots, 2r-2$ and vertical coordinates $2r-1$, $2r, 2r+1, \cdots, 4r-4$. The coordinates of the cells of the lattice diagram give all conceivably possible number-pairs which might occur in the b-, c-, $\cdots$, and

	1	2	3	4	5	6	$\cdots$	$2r-5$	$2r-4$	$2r-3$	$2r-2$
$2r-1$	P	X	X		X		$\cdots$	X		X	
$2r$	X	P		X		X	$\cdots$		X		X
$2r+1$	X		P	X	X		$\cdots$	X		X	
$2r+2$		X	X	P		X	$\cdots$		X		X
$2r+3$	X		X		P	X	$\cdots$	X		X	
$2r+4$		X		X	X	P	$\cdots$		X		X
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$4r-7$	X		X		X		$\cdots$	P	X	X	
$4r-6$		X		X		X	$\cdots$	X	P		X
$4r-5$	X		X		X		$\cdots$	X		P	X
$4r-4$		X		X		X	$\cdots$		X	X	P

DIAGRAM 1

l-groups. The cells whose coordinates appear as a pair in some block of the design are marked with a P and the cells whose coordinates are ruled out by Lemma 5.1 or one of its corollaries are marked with a X. The cells corresponding to the number-pairs appearing in the b_1-set are those lying in the upper left corner of the subsquares of side 2 lying along the main diagonal of the lattice diagram.

The number-pairs whose coordinates are the numbers occurring with i in the a_1- and b_1-sets, where $i = 1, 3, 5, \cdots, 2r - 3$, are ruled out by Lemma 5.1. So are the number-pairs whose coordinates are the numbers occurring with j in the a_2- and b_1-sets, where $j = 2r - 1, 2r + 1, 2r + 3, \cdots, 4r - 5$. The cells corresponding to these number-pairs lie in the upper right and lower left corners of the subsquares of side 2 lying along the main diagonal of the lattice diagram. Furthermore, Lemma 5.1 rules out the occurrence of the number-pairs indicated in Diagram 2.

	1	3	5	$\cdots$	$2r - 5$	$2r - 3$
$2r - 1$	---	$(3, 2r - 1)$	$(5, 2r - 1)$	$\cdots$	$(2r - 5, 2r - 1)$	$(2r - 3, 2r - 1)$
$2r + 1$	$(1, 2r + 1)$	---	$(5, 2r + 1)$	$\cdots$	$(2r - 5, 2r + 1)$	$(2r - 3, 2r + 1)$
$2r + 3$	$(1, 2r + 3)$	$(3, 2r + 3)$	---	$\cdots$	$(2r - 5, 2r + 3)$	$(2r - 3, 2r + 3)$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$4r - 7$	$(1, 4r - 7)$	$(3, 4r - 7)$	$(5, 4r - 7)$	$\cdots$	---	$(2r - 3, 4r - 7)$
$4r - 5$	$(1, 4r - 5)$	$(3, 4r - 5)$	$(5, 4r - 5)$	$\cdots$	$(2r - 5, 4r - 5)$	---

Diagram 2

The corresponding cells in the lattice (Diagram 1) are those lying in the odd numbered columns and odd numbered rows, with the exception of those lying along the main diagonal of the lattice diagram. In other words, they are all cells lying in the upper left corners of the subsquares of side 2, with the exception of those lying along the main diagonal of the lattice diagram.

Treatments $2r - 1, 2r + 1, 2r + 3, \cdots, 4r - 5$ must each occur in each of the remaining $r - 2$ groups (Corollary 5.2). The available number-pairs involving these treatments are given in Diagram 3. Since there are only $r - 2$ number-

	2	4	6	$\cdots$	$2r - 4$	$2r - 2$
$2r - 1$	---	$(4, 2r - 1)$	$(6, 2r - 1)$	$\cdots$	$(2r - 4, 2r - 1)$	$(2r - 2, 2r - 1)$
$2r + 1$	$(2, 2r + 1)$	---	$(6, 2r + 1)$	$\cdots$	$(2r - 4, 2r + 1)$	$(2r - 2, 2r + 1)$
$2r + 3$	$(2, 2r + 3)$	$(4, 2r + 3)$	---	$\cdots$	$(2r - 4, 2r + 3)$	$(2r - 2, 2r + 3)$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$4r - 7$	$(2, 4r - 7)$	$(4, 4r - 7)$	$(6, 4r - 7)$	$\cdots$	---	$(2r - 2, 4r - 7)$
$4r - 5$	$(2, 4r - 5)$	$(4, 4r - 5)$	$(6, 4r - 5)$	$\cdots$	$(2r - 4, 4r - 5)$	---

Diagram 3

pairs involving each of the treatments $2r - 1, 2r + 1, 2r + 3, \cdots$, and $4r - 5$, each of the number-pairs in the Diagram 3 must occur in the design.

Comparing the available number-pairs of Diagram 3 with those in the a_2-set, and using Lemma 5.1, rules out the number-pairs indicated in Diagram 4. Thus

	2	4	6	$\cdots$	$2r-4$	$2r-2$
$2r$	---	$(4, 2r)$	$(6, 2r)$	$\cdots$	$(2r-4, 2r)$	$(2r-2, 2r)$
$2r+2$	$(2, 2r+2)$	---	$(6, 2r+2)$	$\cdots$	$(2r-4, 2r+2)$	$(2r-2, 2r+2)$
$2r+4$	$(2, 2r+4)$	$(4, 2r+4)$	---	$\cdots$	$(2r-4, 2r+4)$	$(2r-2, 2r+4)$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$4r-6$	$(2, 4r-6)$	$(4, 4r-6)$	$(6, 4r-6)$	$\cdots$	---	$(2r-2, 4r-6)$
$4r-4$	$(2, 4r-4)$	$(4, 4r-4)$	$(6, 4r-4)$	$\cdots$	$(2r-4, 4r-4)$	---

DIAGRAM 4

in the lattice (Diagram 1) we cross out the cells lying in the lower right corners of all subsquares of side 2, except for those lying along the main diagonal of the lattice diagram.

Examination of the lattice diagram now shows that the only cells for which both coordinates are even numbers are those lying along the main diagonal, that is, cells whose coordinates are

$$(2, 2r), \quad (4, 2r+2), \quad \cdots, \quad (2r-2, 4r-4).$$

Thus the pairing off of the numbered treatments appearing in the b_2-set is uniquely determined. The $r-1$ blocks of the b_2-set are:

$$\begin{array}{lll} b_2 & 2 & 2r \\ b_2 & 4 & 2r+2 \\ b_2 & 6 & 2r+4 \\ \cdots & \cdots & \cdots \\ b_2 & 2r-2 & 4r-4. \end{array}$$

The occurrence of the above number-pairs in the b_2-set rules out only number-pairs which have been previously excluded.

Each of the treatments $2r, 2r+2, 2r+4, \cdots, 4r-4$ must occur in each of the remaining $r-2$ groups (Corollary 5.2). The only available number-pairs involving these treatments are given in Diagram 5. Again, there are only $r-2$ pairs involving the treatments in question. Hence, each of the number-pairs in

	1	3	5	$\cdots$	$2r-5$	$2r-3$
$2r$	---	$(3, 2r)$	$(5, 2r)$	$\cdots$	$(2r-5, 2r)$	$(2r-3, 2r)$
$2r+2$	$(1, 2r+2)$	---	$(5, 2r+2)$	$\cdots$	$(2r-5, 2r+2)$	$(2r-3, 2r+2)$
$2r+4$	$(1, 2r+4)$	$(3, 2r+4)$	---	$\cdots$	$(2r-5, 2r+4)$	$(2r-3, 2r+4)$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$4r-6$	$(1, 4r-6)$	$(3, 4r-6)$	$(5, 4r-6)$	$\cdots$	---	$(2r-3, 4r-6)$
$4r-4$	$(1, 4r-4)$	$(3, 4r-4)$	$(5, 4r-4)$	$\cdots$	$(2r-5, 4r-4)$	---

DIAGRAM 5

Diagram 5 must occur in the design. Furthermore, these number-pairs and those in Diagram 3 involving treatments $2r - 1, 2r + 1, 2r + 3, \cdots, 4r - 5$ are the only available number-pairs.

In each of the number-pair diagrams of available pairs, it is obvious that number-pairs lying in the same row or in the same column are not set compatible. Furthermore, the occurrence in any set of two unsymmetrically located number-pairs with respect to the main diagonal of the number-pair diagram would rule out the occurrence of the number-pair lying at the interesection of the rows and columns containing the two number-pairs in question (Corollary 5.5). Hence, only symmetrically located pairs with respect to the main diagonal of the number-pair diagram can possibly be set compatible.

Now each set contains $r - 1$ blocks and consequently requires $r - 1$ number-pairs. But we have available a maximum of 4 possibly set-compatible pairs, two from each number-pair diagram. Hence, all designs of the series (7.1) having $r \geqq 6$ are combinatorially impossible.

We shall now give a construction for the design of (7.1) corresponding to $r = 5$. Its parameters are

$$(7.3)\qquad \begin{cases} v = 27, & r = 5, \quad \lambda_1 = 1, \quad n_1 = 10, \\ b = 45, & k = 3, \quad \lambda_2 = 0, \quad n_2 = 16, \\ & P_1 = \begin{pmatrix} 1 & 8 \\ 8 & 8 \end{pmatrix}, \quad P_2 = \begin{pmatrix} 5 & 5 \\ 5 & 10 \end{pmatrix}. \end{cases}$$

By the preceding argument it is seen that we may write, without loss of generality, the five θ-blocks, the eight mixed blocks of the a-group, the eight blocks of the b-group, and the distribution of the lettered treatments throughout the remaining mixed blocks as shown below. The placing of the underlined numbered treatments comes later.

θ	a_1	a_2
θ	b_1	b_2
θ	c_1	c_2
θ	d_1	d_2
θ	e_1	e_2

a-Group			b-Group			c-Group			d-Group			e-Group		
a_1	1	2	b_1	1	9	c_1	$\underline{4}$	$\underline{9}$	d_1	$\underline{6}$	$\underline{9}$	e_1	$\underline{8}$	$\underline{9}$
a_1	3	4	b_1	3	11	c_1	$\underline{2}$	$\underline{11}$	d_1	$\underline{2}$	$\underline{13}$	e_1	$\underline{2}$	$\underline{15}$
a_1	5	6	b_1	5	13	c_1	$\underline{7}$	$\underline{14}$	d_1	$\underline{7}$	$\underline{12}$	e_1	$\underline{5}$	$\underline{12}$
a_1	7	8	b_1	7	15	c_1	$\underline{5}$	$\underline{16}$	d_1	$\underline{3}$	$\underline{16}$	e_1	$\underline{3}$	$\underline{14}$
a_2	9	10	b_2	2	10	c_2	$\underline{8}$	$\underline{13}$	d_2	$\underline{8}$	$\underline{11}$	e_2	$\underline{6}$	$\underline{11}$
a_2	11	12	b_2	4	12	c_2	$\underline{6}$	$\underline{15}$	d_2	$\underline{4}$	$\underline{15}$	e_2	$\underline{4}$	$\underline{13}$
a_2	13	14	b_2	6	14	c_2	$\underline{3}$	$\underline{10}$	d_2	$\underline{5}$	$\underline{10}$	e_2	$\underline{7}$	$\underline{10}$
a_2	15	16	b_2	8	16	c_2	$\underline{1}$	$\underline{12}$	d_2	$\underline{1}$	$\underline{14}$	e_2	$\underline{1}$	$\underline{16}$

DIAGRAM 6.—Plan for design (7.3)

From Diagram 3 of available number-pairs it is seen that the only available number-pairs involving treatments 2, 4, 6, and 8 are those shown in Diagram 7.

	2	4	6	8
9	---	(4, 9)	(6, 9)	(8, 9)
11	(2, 11)	---	(6, 11)	(8, 11)
13	(2, 13)	(4, 13)	---	(8, 13)
15	(2, 15)	(4, 15)	(6, 15)	---

DIAGRAM 7

From Diagram 5 of available number-pairs we also see that the only available number-pairs involving treatments 1, 3, 5, and 7 are those shown in Diagram 8.

	1	3	5	7
10	---	(3, 10)	(5, 10)	(7, 10)
12	(1, 12)	---	(5, 12)	(7, 12)
14	(1, 14)	(3, 14)	---	(7, 14)
16	(1, 16)	(3, 16)	(5, 16)	---

DIAGRAM 8

Since these are the only available number-pairs and since each of the numbered treatments must occur in each of the *c*-, *d*-, and *e*-groups, by Corollary 5.2 the design must contain all the number-pairs in these two arrays (Diagrams 7 and 8).

Now, by Corollary 5.2 the number-pairs (4, 9), (6, 9), and (8, 9) must be distributed one in each of the last three groups. Suppose we arbitrarily form the three blocks

$$\begin{array}{ccc} c_1 & 4 & 9 \\ d_1 & 6 & 9 \\ e_1 & 8 & 9. \end{array}$$

Since each set must contain four number-pairs and since only symmetrically located pairs with respect to the main diagonal of the number-pair diagram are set compatible, we must also form the blocks:

$$\begin{array}{ccc} c_1 & 2 & 11 \\ d_1 & 2 & 13 \\ e_1 & 2 & 15. \end{array}$$

The occurrence of the pairs (4, 9) and (2, 11) in the c_1-set precludes, by Corollary 5.5, the occurrence in this set of pairs involving 1, 3, 10, and 12. Thus, the c_1-set must be completed with (5, 16) and (7, 14). Then by Corollary 5.2, we must complete the c_2-set by use of pairs (8, 13), (6, 15), (3, 10) and (1, 12).

The occurrence of pairs (6, 9) and (2, 13) in the d_1-set precludes, by Corollary 5.5, the occurrence in the same set of any pairs involving 1, 4, 5, 8, 10, 11, 14, and 15. Thus, it is seen that the only pairs which can occur with (6, 9) and (2, 13) in the d_1-set are (7, 12) and (3, 16). Then by Corollary 5.2 the d_2-set must contain pairs (8, 11), (4, 15), (5, 10), and (1, 14).

Since the e_1-set contains pairs (8, 9) and (2, 15), Corollary 5.5 rules out the occurrence in the e_1-set of any pair involving 1, 7, 10, 16, 4, 6, 11, or 13. Thus, it is seen that only (5, 12) and (3, 14) are available. Then by Corollary 5.2 the pairs (6, 11), (4, 13), (7, 10), and (1, 16) must occur in the e_2-set. This completes the construction of design (7.3). The conditions set forth in Lemma 5.1 and in Corollaries 5.1 through 5.5 are satisfied, and Theorems 3.1 and 3.2 are also satisfied.

When $r = 2$, the design (7.1) is a lattice design obtained by arranging 9 treatments in a 3×3 square and taking the rows and columns for blocks. The corresponding design of (7.2) is the dual of this and is a group divisible design [3] with 6 treatments divided into two groups, say 1, 3, 5 and 2, 4, 6. The 9 blocks are obtained by taking all possible pairs consisting of one treatment from each group.

When $r = 3$, both series (7.1) and (7.2) lead to the same symmetrical (self-dual) design with 15 treatments and 15 blocks. This is a triangular design with known solution [3]. Hence we may state the following theorem:

THEOREM 7.1. *The series of designs* (7.1) *contains only three combinatorially possible designs:*

(1) *the lattice design with* $v = 9$, $r = 2$, and $k = 3$;

(2) *the symmetrical triangular design with* $v = b = 15$, $r = k = 3$, $n_1 = 6$, $n_2 = 8$, $\lambda_1 = 1$, $\lambda_2 = 0$;

(3) *the design with parameters* (7.3) *and plan following* (7.3).

Combining the results of this section with those of Sections 4, 5, and 6, we may state the following theorem:

THEOREM 7.2. *Partially balanced incomplete block designs with two associate classes for which* $k > r = 3$, *with* $\lambda_1 = 1$ *and* $\lambda_2 = 0$, *must belong to one of the following classes:*

(a) *Designs obtained by dualizing balanced incomplete block designs with* $k = 3$ *and* $\lambda = 1$. *These designs belong to series* (4.34).

(b) *Lattice designs with three replications belonging to series* (4.33).

(c) *The design of series* (4.32) *for which* $k = 5$. *This is the dual of the design with parameters* (7.3).

REFERENCES

[1] R. C. BOSE, "Partially balanced incomplete block designs with two associate classes involving only two complete replications," *Bull. Calcutta Stat. Assn.*, Vol. 3 (1951), pp. 120–125.

[2] R. C. BOSE AND K. P. NAIR, "Partially balanced incomplete block designs," *Sankhyā*, Vol. 4 (1939), pp. 337-372.

[3] R. C. Bose and T. Shimamoto, "Classification and analysis of partially balanced incomplete block designs with two associate classes," *J. Amer. Stat. Assn.*, Vol. 47 (1952), pp. 151–184.

[4] W. S. Connor and W. H. Clatworthy, "Some theorems for partially balanced designs," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 100–112.

[5] K. R. Nair and C. R. Rao, "A note on partially balanced incomplete block designs," *Science and Culture*, Vol. 7 (1942), pp. 568–569.

[6] K. R. Nair, "Certain inequality relationships among the combinatorial parameters of incomplete block designs," *Sankhyā*, Vol. 6 (1943), pp. 255–259.

[7] M. Reiss, "Ueber eine steinerche combinatorische Aufgabe," *J. Reine Angew. Math.*, Vol. 56 (1895), pp. 326–344.

[8] S. S. Shrikhande, "On the dual of some balanced incomplete block designs," *Biometrics*, Vol. 8 (1952), pp. 66–72.

THE DISTRIBUTION OF LENGTH AND COMPONENTS OF THE SUM OF n RANDOM UNIT VECTORS[1]

By J. Arthur Greenwood and David Durand

Manhattan Life Insurance Company and Massachusetts Institute of Technology

0. Summary. Statistical problems involving angular observations may arise in diverse scientific fields, either from direct measurement of angles—say the direction of winds or of glacial pebbles or of fracture planes—or they may arise from the measurement of times reduced modulo some period and converted into angles—say time of day when train wrecks occur. Specifically, we consider a set of n points ξ_ν situated on a unit circle and assumed to constitute a sample from a distribution having the p.d.f. $g(\xi)$, where $0 \leqq \xi < 2\pi$. Let the n random unit vectors thus defined have the components $\sin \xi_\nu$ and $\cos \xi_\nu$, and set

$$(0.1) \qquad V = \sum \cos \xi_\nu, \qquad W = \sum \sin \xi_\nu, \qquad R = \sqrt{V^2 + W^2}.$$

Let $P(r, n)$ be the probability that $R \leqq r$, and let $Q = 1 - P(r, n)$.

This paper shows how the statistics V and R provide tests for the uniform distribution $g(\xi) = 1/2\pi$. The distribution of R on the hypothesis of uniformity was derived by Kluyver as a solution to Pearson's random walk problem, and is tabulated here for use in significance tests. The distribution of V is derived here, but has not been calculated.

To illustrate the type of tests that might employ the statistics R and V, consider a carnival wheel—first from the standpoint of the punter who suspects bias, second from the standpoint of the mechanic who has attempted to introduce bias. The punter, by studying the performance of the wheel, might wish to answer two questions: first, does the wheel differ credibly from an unbiased wheel? second, what is the direction and extent of the bias, if any? An answer to the first question is obtainable from the distribution of R, and an answer to the second has been provided by Mises. The mechanic, on the other hand—because he knows the direction of the bias, if there is a bias—might better use the statistic V as a test of his success, and he might appropriately modify the Mises approach in estimating the extent of the bias.

1. Pearson's random walk. In 1905 Pearson [13] posed the following problem: "A man starts from a point O and walks a distance a in a straight line; he then turns through any angle whatever and walks a distance a in a second straight line. He repeats this process n times.

Received April 6, 1954.

[1] This is a revised version of a paper presented at the Kingston meeting of the Institute of Mathematical Statistics in September 1953, under the title "The Integral Solution of Pearson's Random Walk Problem and Related Matters."

"I require the probability that after these n stretches he is at a distance between r and $r + \delta r$ from his starting point, O."

Thus, Pearson was assuming a sample of n random angles ξ_ν from the uniform distribution $g(\xi) = 1/2\pi$, and his required probability is the differential.

$$\frac{d}{dr} P(r, n)\, \delta r. \tag{1.1}$$

After Pearson's statement of the random walk problem, Rayleigh [16] observed an analogy to the theory of vibrations and deduced the asymptotic formula

$$Q(r, n) = \exp[-r^2/n]. \tag{1.2}$$

Then Kluyver [7] obtained a solution, showing that the probability of $R \leqq r$ is

$$P(r, n) = r \int_0^\infty [J_0(x)]^n J_1(rx)\, dx. \tag{1.3}$$

This formal result can easily be found by bivariate characteristic functions, but the transformation from rectangular to polar coordinates is difficult to make rigorous.

Both Kluyver and Pearson were loath to attempt direct quadrature of (1.3) or its p.d.f. (see [14], p. 5), and Pearson devoted his efforts toward developing an asymptotic approximation (Section 6). We, however, felt that Pearson's series could best be checked by quadratures and that, with the advent of extended tables of the Bessel functions and improved calculating equipment, this was now feasible. Table 1 presents quadrature values of Kluyver's integral. Table 2 presents interpolated 5 percent and 1 percent points for r, and for several functions of r that may be useful in making significance tests. A discussion of the calculation procedure appears in Sections 4 and 5.

2. The problem of Mises. In another problem involving the summation of random unit vectors, Mises [12] considered vectors distributed according to the p.d.f.

$$g(\xi) = \frac{\exp\{k \cos(\xi - \alpha)\}}{2\pi I_0(k)}, \tag{2.1}$$

of which the uniform distribution is a degenerate case ($k = 0$). He then showed that a joint maximum likelihood estimate of the concentration parameter k and the modal angle α is obtainable from the vector sum by means of the relations

$$R \cos \alpha = \sum \cos \xi_\nu = V, \qquad R \sin \alpha = \sum \sin \xi_\nu = W. \tag{2.2}$$

Recently Gumbel, Greenwood, and Durand [6] christened (2.1) the "circular normal distribution," tabulated the integral thereof, and calculated a table for converting $R/n = \bar{a}$ into $\hat{k}$, the maximum likelihood estimate[2] of k.

[2] According to Mises ([12], eq. 16), $R/n = I_1(k)/I_0(k)$, which is tabulated by Gumbel et al. ([6], p. 140).

TABLE 1

$$P(r, n) = r \int_0^\infty [J_0(x)]^n J_1(rx)\, dx$$ *(See Section 5 before attempting to interpolate in this table)*

r	n = 6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
0.5	.03885	.03229	.02911	.02588	.02353	.02148	.01979	.01834	.01709	.01599	.01503	.01418	.01342	.01274	.01212	.01156	.01105	.01058	.01015
1.0	.14286	.12500	.11111	.10000	.09091	.08333	.07692	.07143	.06667	.06250	.05882	.05556	.05263	.05000	.04762	.04545	.04348	.04167	.04000
1.5	.29324	.26141	.23345	.21182	.19348	.17821	.16511	.15383	.14398	.13533	.12765	.12080	.11464	.10908	.10404	.09944	.09523	.09137	.08780
2.0	.46782	.41782	.37907	.34622	.31884	.29535	.27512	.25746	.24193	.22816	.21587	.20483	.19487	.18583	.17759	.17005	.16312	.15673	.15083
2.5	.63165	.57455	.52795	.48750	.45291	.42271	.39626	.37287	.35207	.33344	.31666	.30149	.28769	.27509	.26354	.25292	.24312	.23405	.22563
3.0	.76722	.71404	.66424	.62132	.58281	.54868	.51809	.49063	.46583	.44336	.42290	.40421	.38708	.37131	.35676	.34330	.33080	.31917	.30833
3.5	.87020	.82278	.77819	.73679	.69872	.66381	.63182	.60249	.57555	.55077	.52791	.50679	.48723	.46906	.45216	.43640	.42168	.40789	.39496
4.0	.93755	.90039	.86465	.82871	.79452	.76189	.73117	.70230	.67524	.64989	.62616	.60393	.58309	.56353	.54516	.52788	.51160	.49625]	.48176
4.5	.97321	.95066	.92400	.89635	.86805	.84009	.81276	.78635	.76099	.73675	.71364	.69165	.67074	.65088	.63201	.61409	.59706	.58086	.56545
5.0	.99121	.97864	.96149	.94198	.92067	.89848	.87590	.85333	.83105	.80923	.78800	.76743	.74756	.72842	.71001	.69231	.67533	.65902	.64337
5.5	.99858	.99205	.98280	.97022	.95562	.93932	.92196	.90389	.88547	.86694	.84847	.83020	.81224	.79466	.77751	.76081	.74459	.72886	.71362
6.0	1.00000	.99788	.99329	.98626	.97705	.96604	.95359	.94005	.92570	.91080	.89555	.88012	.86463	.84920	.83390	.81879	.80393	.78934	.77506
6.5		.99976	.99783	.99442	.98916	.98233	.97403	.96451	.95397	.94263	.93067	.91824	.90549	.89251	.87942	.86629	.85318	.84014	.82723
7.0		1.00000	.99953	.99804	.99542	.99151	.98640	.98014	.97285	.96468	.95576	.94621	.93616	.92570	.91494	.90395	.89281	.88157	.87029
7.5			.99996	.99945	.99830	.99629	.99338	.98954	.98451	.97924	.97291	.96591	.95833	.95025	.94174	.93289	.92375	.91439	.90486
8.0			1.00000	.99990	.99946	.99855	.99704	.99485	.99197	.98838	.98412	.97923	.97376	.96776	.96130	.95442	.94719	.93965	.93185
8.5				1.00000	.99987	.99951	.99880	.99765	.99601	.99383	.99111	.98786	.98408	.97981	.97508	.96994	.96441	.95853	.95235
9.0					.99998	.99986	.99957	.99902	.99815	.99691	.99527	.99320	.99071	.98780	.98448	.98076	.97668	.97225	.96751
9.5					1.00000	.99997	.99986	.99963	.99920	.99855	.99761	.99637	.99480	.99290	.99065	.98807	.98516	.98193	.97840
10.0						1.00000	.99997	.99988	.99969	.99936	.99886	.99816	.99722	.99603	.99457	.99234	.99084	.98856	.98601
10.5							.99999	.99996	.99989	.99974	.99949	.99911	.99858	.99787	.99696	.99585	.99452	.99296	.99118
11.0							1.00000	.99999	.99997	.99990	.99979	.99960	.99931	.99890	.99836	.99768	.99682	.99580	.99460
11.5								1.00000	.99999	.99997	.99992	.99983	.99968	.99946	.99916	.99875	.99822	.99757	.99678
12.0									1.00000	.99999	.99997	.99993	.99986	.99975	.99958	.99935	.99904	.99864	.99814
13.0										1.00000	1.00000	.99999	.99998	.99995	.99991	.99985	.99975	.99962	.99944
14.0												1.00000	1.00000	.99999	.99999	.99997	.99995	.99991	.99985
15.0														1.00000	1.00000	1.00000	.99999	.99998	.99997
16.0																	1.00000	1.00000	.99999
17.0																			1.00000

TABLE 2

5 and 1 percent points for r and derived functions

n	5 percent — $P(r, n) = .95$				1 percent — $P(r, n) = .99$			
	r	r/n	r^2	$z = r^2/n$	r	r/n	r^2	$z = r^2/n$
6	4.141	.6901	17.14	2.8573	4.951	.8251	24.51	4.0851
7	4.491	.6416	20.17	2.8819	5.394	.7705	29.09	4.1561
8	4.818	.6022	23.21	2.9014	5.797	.7246	33.61	4.2007
9	5.118	.5686	26.19	2.9102	6.185	.6872	38.25	4.2504
10	5.402	.5402	29.19	2.9187	6.550	.6550	42.90	4.2899
11	5.674	.5158	32.19	2.9262	6.893	.6266	47.51	4.3195
12	5.932	.4943	35.18	2.9320	7.220	.6017	52.13	4.3441
13	6.179	.4753	38.18	2.9370	7.533	.5795	56.75	4.3651
14	6.417	.4584	41.18	2.9413	7.833	.5595	61.36	4.3829
15	6.646	.4431	44.18	2.9450	8.122	.5415	65.97	4.3983
16	6.868	.4293	47.17	2.9482	8.402	.5251	70.59	4.4116
17	7.083	.4166	50.17	2.9511	8.672	.5101	75.20	4.4234
18	7.291	.4051	53.16	2.9536	8.934	.4963	79.81	4.4338
19	7.494	.3944	56.16	2.9558	9.188	.4836	84.42	4.4430
20	7.691	.3846	59.16	2.9579	9.435	.4718	89.03	4.4514
21	7.884	.3754	62.15	2.9597	9.677	.4608	93.64	4.4589
22	8.072	.3669	65.15	2.9613	9.912	.4505	98.24	4.4657
23	8.255	.3589	68.15	2.9629	10.142	.4409	102.85	4.4719
24	8.435	.3514	71.14	2.9642	10.366	.4319	107.46	4.4775
∞				2.9957				4.6052

Accordingly, if the punter wishes a joint maximum likelihood estimate of the direction and extent of bias in the carnival wheel, he may easily obtain this if he is willing to assume that the wheel is biased according to the Mises distribution law. Moreover, if he wishes to test whether the observed performance of the wheel is consistent with the uniform hypothesis $k = 0$, he may do so by a simple extension of Mises' argument. In fact, the likelihood ratio for testing $k = 0$ against the alternative $k = \hat{k}$ and $\alpha = \hat{\alpha}$ is

$$\frac{(2\pi)^{-n}}{[2\pi\, I_0(\hat{k})]^{-n} \exp[\hat{k}\Sigma \cos(\xi_\nu - \hat{\alpha})]} = \frac{[I_0(\hat{k})]^n}{\exp[\hat{k}R]}.$$

From the fact that the second derivative of $\ln I_0(k)$ is the variance of the linear distribution of $\cos \xi$, and therefore essentially positive, it can be shown that the above ratio, for fixed n, is a decreasing function of R alone, where R has the distribution (1.3) on $k = 0$. The test of $k = 0$ then consists in comparing R or R^2 with the tabled values of r or r^2 (Table 2). An example will be given in Section 8.

The above hypothesis $k = \hat{k}$ and $\alpha = \hat{\alpha}$—the punter's alternative—is but one of several against which uniformity might be tested. In fact, twelve possible

hypotheses can easily be formed by coupling one of the three α-hypotheses with one of the four k-hypotheses:

$$0 \leqq \alpha < 2\pi, \qquad \alpha_1 \leqq \alpha \leqq \alpha_2, \qquad \alpha_1 = \alpha;$$

$$k = k^* > 0, \ (k^* \text{ unknown}), \qquad k = k_1, \qquad k \geqq k_1, \qquad k_2 \geqq k \geqq k_1.$$

In particular, $\alpha = \alpha_1$ and $k = k^* > 0$ is the appropriate alternative for the mechanic who has attempted to bias the carnival wheel; the mechanic, unlike the punter, knows the position of the mode if there is a mode. And if he rejects uniformity, the mechanic needs to estimate only the parameter k.

The mechanic's problem is of special interest because of its relation to the work of Mises, who formulated his "Kriterium der Ganzzahligkeit" ([12], p. 495 ff.) in hopes of showing that atomic weights are integers subject to error. The observed atomic weights reduced modulo 1 and transformed into angles are subject to a test of the uniformity hypothesis against the alternative $\alpha = 0$ and $k = k^* > 0$. In making this test, Mises proceeded to estimate k from the statistic R; but, on the alternative stated, V and not R furnishes a maximum likelihood estimate of k.

The likelihood $\ln L = -n \ln I_0(k) + kV - n \ln 2\pi$. Setting

$$\frac{d}{dk} \ln L = -n \frac{I_1(k)}{I_0(k)} + V = 0,$$

we find that $V/n = I_1(k)/I_0(k)$, so that the recent table ([6], p. 140) applies. The distribution of V is derived from the characteristic function of $\cos \xi$,

$$\phi(x) = \frac{1}{2\pi} \int_0^{2\pi} e^{ix\cos\xi} \, d\xi = J_0(x).$$

Therefore, the sum of n independent cosines has the characteristic function $[J_0(x)]^n$ and the distribution function

$$C + \frac{1}{2\pi} \int_{-\infty}^{\infty} [J_0(x)]^n \frac{1 - e^{-ixv}}{ix} \, dx = C + \frac{1}{\pi} \int_0^{\infty} [J_0(x)]^n \frac{\sin vx}{x} \, dx.$$

When $v = 0$, the integral vanishes; therefore $C = \frac{1}{2}$ and the probability that $V \leqq v$ takes the form

$$\frac{1}{2} + \frac{1}{\pi} \int_0^{\infty} [J_0(x)]^n \frac{\sin vx}{x} \, dx \tag{2.3}$$

on the hypothesis $k = 0$. In the limit, V is normally distributed with mean zero and variance $n/2$.

3. Nonzero k and the power function. Under some circumstances the mechanic may feel sanguine that he has effectively biased the wheel and yet have doubts whether the direction of bias is exactly zero, as designed. Then, an entirely different sort of test would be needed—in fact one analogous to the

linear t-test for the difference between an observed mean and its hypothetical value. But the construction of this test—and several others that might be useful—requires investigation of the distribution of R or of $\hat{\alpha} - \alpha$, either jointly or severally, on the hypothesis $k = k_1 > 0$.

On this hypothesis the joint distribution of V and W (or of R and $\hat{\alpha}$), the distribution of R, and the distribution of V can be obtained by an artifice due to Fisher[3] [4], which derives these distributions painlessly from those on the hypothesis $k = 0$. Fisher seems to effect his derivation by means of the following

LEMMA. *Let $F(x, y)$ and $G(x, y)$ be two distribution functions such that*

$$dF(x,y) = Ae^{ax+by}\, dG(x, y), \qquad A^{-1} = \iint e^{ax+by}\, dG(x, y). \tag{3.1}$$

Denote the distributions of the n-fold convolutions of $F(x, y)$ and $G(x, y)$ respectively by $F_n(v, w)$ and $G_n(v, w)$. Then

$$dF_n(v, w) = A^n e^{av+bw}\, dG_n(v, w). \tag{3.2}$$

To prove this let

$$\phi(t, u) = \iint e^{i(tx+uy)}\, dF(x, y),$$

$$\psi(t, u) = \iint e^{i(tx+uy)}\, dG(x, y),$$

where the integrals are Stieltjes integrals extended over the x, y-plane. Then

$$\phi^n(t, u) = \iint e^{i(tv+uw)}\, dF_n(v, w),$$

$$\psi^n(t, u) = \iint e^{i(tv+uw)}\, dG_n(v, w).$$

By hypothesis (3.1)

$$\begin{aligned}\phi(t, u) &= \iint e^{i(tx+uy)}\, Ae^{ax+by}\, dG(x, y)\\ &= A\psi[(t - ia), (u - ib)].\end{aligned}$$

The characteristic functions of the left and right sides of (3.2) are, respectively,

$$\phi^n(t, u) = A^n\psi^n[(t - ia), (u - ib)];$$

$$\iint e^{i(tv+uw)} A^n e^{av+bw}\, dG_n(v, w) = A^n\psi^n[(t - ia), (u - ib)].$$

The equality (3.2) follows from the equality of the characteristic functions.

[3] Fisher's work is concerned with the spherical analogue of the Mises distribution (2.1). Although the trigonometry is more involved in three dimensions than in two, the algebra is much simpler. See Rayleigh [17], p. 338 ff.

On the hypothesis $k = 0$, the differential of (2.3) is

$$dP = \frac{dv}{\pi} \int_0^\infty [J_0(x)]^n \cos vx \, dx.$$

Therefore, by (3.2) the differential of probability on the hypothesis $k = k_1$ and $a = 0$ is

$$dP = \frac{e^{k_1 v} \, dv}{2\pi [I_0(k_1)]^n} \int_{-\infty}^{\infty} [J_0(x)]^n e^{-ivx} \, dx,$$

so that

$$P(V \leqq v) = \frac{1}{2\pi [I_0(k_1)]^n} \int_{-\infty}^{\infty} [J_0(x)]^n \frac{e^{(k_1 - ix)v}}{k_1 - ix} \, dx.$$

This distribution is the power function of V against the Mises alternative $\alpha = 0$ and $k = k_1 > 0$.

To obtain the distributions of W and R, one first writes down the differential of the joint probability of V and W on the hypothesis $k = 0$. The probability density function corresponding to (1.3) is circularly symmetric and has the form

$$\frac{1}{2\pi r} \frac{d}{dr} r \int_0^\infty [J_0(x)]^n J_1(rx) \, dx = \frac{1}{2\pi} \int_0^\infty [J_0(x)]^n J_0(rx) \, x \, dx, \tag{3.4}$$

so that

$$dP(V \leqq v, \quad W \leqq w) = \frac{dv \, dw}{2\pi} \int_0^\infty [J_0(x)]^n J_0(x\sqrt{v^2 + w^2}) x \, dx.$$

It will be convenient to use Pearson's notation of $\phi_n(v^2 + w^2)$ for the integral above. Then, on the hypothesis $\alpha = 0$ and $k = k_1 > 0$

$$dP(V \leqq v, \quad W \leqq w) = (2\pi)^{-1} [I_0(k_1)]^{-n} e^{k_1 v} \phi_n(v^2 + w^2) \, dv \, dw. \tag{3.5}$$

This differential does not readily yield the marginal distribution of W. Probably a convenient approximation is to substitute Pearson's Laguerre function expansion for $\phi_n(v^2 + w^2)$ and integrate v out (see Section 6).

The distribution of R can be obtained by transforming (3.5) into polar coordinates,

$$dP(V \leqq r \cos \beta, \quad W \leqq r \sin \beta) = \frac{e^{k_1 r \cos \beta} \phi_n(r^2)}{2\pi [I_0(k_1)]^n} r \, dr \, d\beta; \tag{3.6}$$

whence $dP(R \leqq r) = [I_0(k_1)]^{-n} I_0(k_1 r) \, \phi_n(r^2) \, r \, dr$. Integration by parts then gives

$$P(R \leqq r) = [I_0(k_1)]^{-n} \left[I_0(k_1 r) P(r, n) - \int_0^r k_1 I_1(k_1 s) P(s, n) \, ds \right],$$

where $P(r, n)$ is the function of (1.3). This last distribution is, in fact, the power function of R against the Mises alternative $\alpha = 0$ and $k = k_1 > 0$. It is also the

power function against the Mises alternative $\alpha = \alpha^*$ (unknown) and $k = k_1 > 0$, because R is unaffected by increasing all the ξ_ν by an arbitrary constant. The distribution of $\hat{\alpha}$ can be evaluated after expanding (3.6) in Laguerre functions.

4. Calculation of Kluyver's integral. The integral has been evaluated for $n = 6(1)24$ and $r = 0(.1)6(.5)12(1)n$, and for such other values at intervals of .05 or .1 as were necessary to embrace the 5 percent and 1 percent points of R. Table 1, as published here, is an abridgment. Evaluation of the integral for $n = 3$, 4, or 5 would have required special treatment, such as: (A) tabulating J_1 to values much greater than the present upper tabled limit of 100; (B) replacing J_0 and J_1 by their asymptotic series and then expanding the integral into a series of sine integrals for $n = 3$ or 5, or a series of Fresnel integrals for $n = 4$; (C) an extension by quadratures as performed by Pearson and Blakeman ([14], pp. 16–20), but carried on digitally instead of graphically. For $n = 2$ the distribution of R is simply $(2/\pi)$ arc sin $r/2$.

The integration formula used was the integral of Bessel's interpolation formula carried to 8 differences. For an interval of integration $h = .04$ was used for $r \leqq 10$, and $h = .02$ for $r \geqq 10$; agreement at $r = 10$ was good.

There were three chief sources of error in the integration:

(A) The integration was truncated at $rx = 100$, the upper limit of the J_1 table. This error is larger than that explained in (B) following for $n = 6$, 7, and 8; for $n = 4$ or 5 it would have been prohibitive. For $n = 7$ arguments for J_1 above 100 were supplied.

(B) The integration was often truncated for $|J_0(x)|^n$ less than 10^{-5}. This was by no means systematic, however, since for many of the integrations additional arguments were conveniently included.

(C) The remainder term, involving ninth differences and greater, was neglected.

The values $P(1, n) = 1/(n + 1)$ and $P(n, n) = 1$, known by elementary analysis, furnish a partial check on the accuracy of the integration. In most of the table discrepancies were less than 10^{-6}, and nowhere were they as great as 10^{-5}. We believe that Table 1 is accurate to 5 decimals as given, with the exception of $n = 6$, where an error of 1 in the fifth place can easily have occurred.

5. Interpolation of percentage points. Because of discontinuities in its derivatives with respect to r, Kluyver's integral (1.3) presents problems of interpolation, both direct and inverse. These discontinuities have been discussed by Rayleigh [17] who investigated the behavior of the leading term in the asymptotic expansion of the Bessel function product. He found that the derivatives of order $\frac{1}{2}(n - 2)$ for n even and $\frac{1}{2}(n - 3)$ for n odd are discontinuous[4] at the points $n - 2k$, for $k = 0, 1, 2, 3$, etc. Although the integral

$$\int_0^x [J_0(x)]^n J_1(rx)\, dx,$$

[4] For a discussion of discontinuities in the sum of n rectangular variables, see Cramér [3], p. 245, and Fisher [4], p. 298.

where the upper limit of quadrature X is a large number, has continuous derivatives of all orders, these derivatives become large for the same points $n - 2k$.

For the problem of inversely interpolating the 5 percent and 1 percent points (Table 2), the singularities of (1.3) were not particularly troublesome because none of the desired points happen to fall uncomfortably close to a singularity. Thus, for the interpolation of $Q(r, 6) = .05$, the function was evaluated for the five arguments 4.10, 4.15, 4.20, 4.25, and 4.30, which do not include the singularity $r = 4.00$. Although singularities are less dangerous for large n—since the order of the first discontinuous derivative increases—they were avoided whenever convenient, and in no case did an interpolation use points on both sides of a singularity. Table 2 presents interpolated 5 percent and 1 percent points for r, and for several functions of r that may be useful in making significance tests. The first three of these functions—namely, r, r/n, and r^2—should all be correct to four significant figures, as tabled. The function z, given to five significant figures for comparisons in Section 7, may possibly contain an error of about 2 in the last figure.

6. Pearson's asymptotic expansion. Pearson derived an asymptotic expansion for the probability density function (3.4) corresponding to (1.3). This expansion is, in effect, a series of Laguerre functions

$$f(r^2) = e^{-kr^2} \sum_{i=0}^{\infty} c_i L_i(kr^2), \qquad L_i(x) = e^x \frac{d^i}{dx^i}(x^i e^{-x}) = \sum_{t=0}^{i} \frac{i!\, i!(-x)^t}{t!\, t!(i-t)!}.$$

To derive the c_i, Pearson essentially used $[J_0(x)]^n$ as a moment generating function, thus obtaining the expansion

$$\frac{1}{2\pi} \int_0^\infty [J_0(x)]^n J_0(rx) x \, dx = \frac{1}{\pi n} e^{-z} \sum_{i=0}^{\infty} c_i L_i(z) \tag{6.1}$$

or, with $z = r^2/n$,

$$2\pi r \, dr \frac{1}{2\pi} \int_0^\infty [J_0(x)]^n J_0(rx) x \, dx = dz \, e^{-z} \sum_{i=0}^{\infty} c_i L_i(z). \tag{6.2}$$

Each c_i ([14], p. 9) is of order $n^{-\beta}$, where β is the integral part of $\frac{1}{2}(1 + i)$. Pearson derived the c_i through c_6, which contains terms in n^{-5}.

From the well-known relation

$$\int_z^\infty e^{-z} L_i(z) \, dz = e^{-z}[L_i(z) - iL_{i-1}(z)],$$

the series on the right of (6.2) may be integrated term by term as

$$Q(r, n) = e^{-z}\left\{1 + \sum_{i=0}^{\infty} c_i[L_i(z) - iL_{i-1}(z)]\right\}. \tag{6.3}$$

Rearranging (6.3) in descending powers of n and omitting contributions of order n^{-4} and n^{-5}, which appear in c_5 and c_6, we get

$$Q(r, n) = e^{-z}\left[1 + \frac{1}{2n}\left(z - \frac{z^2}{2!}\right) + \frac{1}{12n^2}\left(-z + \frac{11z^2}{2!} - \frac{19z^3}{3!} + \frac{9z^4}{4!}\right)\right.$$
$$\left. + \frac{1}{24n^3}\left(-2z - \frac{4z^2}{2!} + \frac{69z^3}{3!} - \frac{163z^4}{4!} + \frac{145z^5}{5!} - \frac{45z^6}{6!}\right)\right]. \tag{6.4}$$

This series is more convenient for computation than (6.3). Table 3 compares for $n = 7$ and 14 the value of $P(r, n) = 1 - Q(r, n)$ obtained by quadratures with Rayleigh's approximation $P(r, n) = 1 - e^{-z}$, Pearson's approximation (6.3) through c_6, and the approximation (6.4) through n^{-3}. It appears that either (6.3) or (6.4) gives roughly three-decimal accuracy for $n = 7$, and that, if anything, (6.4) is a little better than (6.3). Although three reliable decimal places provide a fair degree of accuracy over most of the curve, they are clearly insufficient for estimating percentage points in the extreme tail.

7. Series expansion of the percentage points. Having obtained $Q(r, n)$ as an expansion in powers of z and n^{-1}, we now attempt to obtain an expansion for z in terms of $Q(r, n)$ and n^{-1}. Since the Rayleigh approximation above obviously leads to $z = -\ln Q(r, n)$, it is convenient to set

$$y = -\ln Q(r, n) \tag{7.1}$$

and expand z in powers of y. To do this, set

$$z = y + \sum a_i n^{-i}. \tag{7.2}$$

Then $e^z Q(r, n) = \exp \sum a_i n^{-i}$, whence we obtain

$$a_1 = \tfrac{1}{4}(2y - y^2), \qquad a_2 = \tfrac{1}{72}(12y - 3y^2 - y^3),$$
$$a_3 = \tfrac{1}{288}(-12y + 42y^2 - 8y^3 - y^4).$$

For $n = 7$, 14, and 21, Table 4 gives the successive approximations to the 1 percent and 5 percent points and compares them with the values obtained by inverse interpolation. Even when four terms (through a_3) are used, good accuracy is not achieved for $n = 7$. Three terms give fair three-place accuracy for $n = 14$ and four-place accuracy for $n = 21$, so that three terms suffice to extend Table 2 beyond $n = 24$. However, three terms may be inadequate to obtain points in the extreme tail, such as the 0.1 percent point, if these are desired. For example, $Q(12, 21) = .000{,}648{,}77$ by quadratures. The three-term approximation for z is 6.8590 against the correct value $144/21 = 6.8571$. The fourth term is $-.0015$, and the four-term approximation is therefore 6.8575.

8. An example of the *R*-test. Forrest [5] gives 651 observed times of breakdown due to lightning in the British electrical grid system for the eight years 1940–47. The observations fall into two groups: 588 for the summer months (April–September) and 63 for the winter months (October–March). The summer

TABLE 3

Various calculations of Kluyver's integral for
$n = 7$ and 14

$$P(r, n) = r \int_0^\infty [J_0(x)]^n J_1(rx)\, dx$$

r	$P(r, n)$ by quadrature	$1 - e^{-r^2/n}$ (Rayleigh)	(6.3) through c_8 (Pearson)	(6.4) through n^{-3}
		$n = 7$		
0.5	.03229	.03508	.03284	.03272
1.0	.12500	.13312	.12430	.12500
1.5	.26141	.27480	.26106	.26078
2.0	.41782	.43528	.41819	.41819
2.5	.57455	.59052	.57469	.57497
3.0	.71404	.72355	.71305	.71339
3.5	.82278	.82623	.82279	.82289
4.0	.90039	.89830	.90095	.90082
4.5	.95066	.94458	.95065	.95046
5.0	.97864	.97188	.97857	.97846
5.5	.99205	.98672	.99219	.99219
6.0	.99788	.99416	.99779	.99785
6.5	.99976	.99761	.99962	.99968
7.0	1.00000	.99909	1.00003	1.00006
		$n = 14$		
0.5	.01709	.01770	.01709	.01709
1.0	.06667	.06894	.06668	.06667
1.5	.14398	.14846	.14401	.14399
2.0	.24193	.24853	.24196	.24194
2.5	.35207	.36009	.35208	.35207
3.0	.46583	.47421	.46583	.46583
3.5	.57555	.58314	.57569	.57555
4.0	.67524	.68109	.67521	.67524
4.5	.76099	.76459	.76098	.76099
5.0	.83105	.83232	.83104	.83105
5.5	.88547	.88476	.88548	.88548
6.0	.92570	.92357	.92571	.92570
6.5	.95397	.95109	.95399	.95397
7.0	.97285	.96980	.97286	.97285
7.5	.98481	.98201	.98481	.98481
8.0	.99197	.98966	.99196	.99196
8.5	.99601	.99427	.99600	.99601
9.0	.99815	.99693	.99814	.99815
9.5	.99920	.99841	.99920	.99921
10.0	.99969	.99921	.99969	.99969
11.0	.99997	.99982	.99997	.99997
12.0	1.00000	.99997	1.00000	1.00000

TABLE 4

Various calculations of the 5 and 1 percent points for $n = 7, 14,$ and 21

	5 percent points $P(r, n) = .95$	1 percent points $P(r, n) = .99$
y	+2.9957 3227	+4.6051 7019
a_1	−0.7457 3683	−2.9993 1302
a_2	−0.2480 4699	−1.4725 7372
a_3	+0.1574 8945	−1.3736 8651
$n = 7$		
$y + a_1/7$	2.8892	4.1767
$y + a_1/7 + a_2/49$	2.8841	4.1466
$y + a_1/7 + a_2/49 + a_3/343$	2.8846	4.1426
z, by interpolation	2.8819	4.1561
$n = 14$		
$y + a_1/14$	2.9425	4.3909
$y + a_1/14 + a_2/196$	2.9412	4.3834
$y + a_1/14 + a_2/196 + a_3/2744$	2.9413	4.3829
z, by interpolation	2.9413	4.3829
$n = 21$		
$y + a_1/21$	2.9602	4.4623
$y + a_1/21 + a_2/441$	2.9597	4.4590
$y + a_1/21 + a_2/441 + a_3/9261$	2.9597	4.4589
z, by interpolation	2.9597	4.4589

breakdown times have a rough resemblance to a Mises distribution with a conspicuous mode at about 1700 hours, though the fit is not first-rate, owing to an apparent secondary mode in the early morning. The winter breakdown times are given in the table below, and it is by no means certain that this distribution differs significantly from the uniform.

We propose to test the hypothesis $k = 0$ against the alternative $k = k^* > 0$ and $0 \leqq \alpha < 2\pi$. To this end, we reduce the midpoints of the hourly time intervals to angles (0–1: 7°.5, etc.) and include the corresponding sines and cosines in Forrest's table.

G.M.T.	Freq.	Sines	Cosines	G.M.T.	Freq.	Sines	Cosines
0–1	1	.1305	.9914	12–13	4	−.1305	−.9914
1–2	1	.3827	.9239	13–14	3	−.3827	−.9239
2–3	1	.6088	.7934	14–15	1	−.6088	−.7934
3–4	3	.7934	.6088	15–16	4	−.7934	−.6088
4–5	1	.9239	.3827	16–17	7	−.9239	−.3827
5–6	2	.9914	.1305	17–18	4	−.9914	−.1305
6–7	1	.9914	−.1305	18–19	2	−.9914	.1305
7–8	3	.9239	−.3827	19–20	4	−.9239	.3827
8–9	0	.7934	−.6088	20–21	4	−.7934	.6088
9–10	7	.6088	−.7934	21–22	1	−.6088	.7934
10–11	0	.3827	−.9239	22–23	6	−.3827	.9239
11–12	1	.1305	−.9914	23–24	2	−.1305	.9914

From this layout, it is a simple matter to calculate the statistics

$$W = -13.3393, \qquad V = -3.2652, \qquad R^2 = 188.59.$$

Since the corresponding percentage points for r^2 have not been tabulated for $n > 24$, it is necessary to calculate an approximate value for z after the method of Section 7. From the two-term formula (7.2), specifically $y + a_1/63$, the 5 percent approximation 2.9835 is obtained, which compares with $R^2/n = 2.993$.

9. Unsolved problems. We have indicated in Section 3 how the distribution of $\hat{\alpha} - \alpha$ may be obtained for known k. To test the significance of an observed discrepancy with no a priori information about k is an obvious extension of the problem in linear statistics solved by the t-test. For large n one can replace k by $\hat{k}$ and treat it as known; for small n the appropriate test is unknown.

On the assumption of discrete sample points measured without error, (1.3) is a suitable mathematical model. Forrest's data, to which we applied (1.3), are sparse data grouped at the midpoints of class intervals. We do not know the effect of such grouping on V, W, or R, or on any estimators derived from them.

Although we have shown that R provides a likelihood ratio test for $k = 0$ against the composite Mises alternative $0 \leqq \alpha < 2\pi$, and $k = k^* > 0$, and that V provides a likelihood ratio test against the Mises alternative $\alpha = \alpha_1$, and $k = k^* > 0$, we feel that these results are by no means conclusive because of the existence of other unimodal, symmetrical p.d.f.s. For example, Levy [9], [10], Marcinkiewicz [11], Perrin ([15], p. 20), Wintner [18], [19], and Zernike ([20], pp. 477–478) have discussed wrapped-up normal and Cauchy distributions —that is, the angle ξ is assumed to have the normal or Cauchy distribution but the angles ξ, $\xi \pm 2\pi$, $\xi \pm 4\pi$, $\cdots$, are indistinguishable. Arnold ([1], p. 6) has shown that V and W jointly provide consistent estimates of the parameters of this distribution. Brooks and Carruthers ([2], p. 199) have translated a bivariate normal distribution into polar coordinates with the origin removed from the center of the distribution and have integrated out with respect to the radius.

Arnold and Krumbein have indicated problems wherein angles are identifiable only modulo π. A workable approach is to use the distribution $Ce^{k\cos 2\xi}$ and the statistics

$$V_2 = \sum \cos 2\xi_\nu, \qquad W_2 = \sum \sin 2\xi_\nu, \qquad R_2 = \sqrt{V_2^2 + W_2^2}.$$

This solution is not unique, and the merits of alternative solutions should be studied.

Spherical distribution problems have received some attention, both as generalizations of circular problems and on their own merits. A few examples might be mentioned. Arnold [1] has extended the Mises distribution and the wrapped-up normal to the surface of the sphere. Rayleigh [17] has shown that the spherical analogue of R has for distribution a set of elementary functions abutting with discontinuous derivatives as in Section 5, and Fisher has given an explicit formula for this distribution. The spherical analogue of the V statistic is distributed as the sum of n rectangular variables (Fisher [4], p. 296).

10. Acknowledgments. We are indebted to E. J. Gumbel, who suggested the problem in 1951 during the joint efforts leading to Gumbel et al. [6], and to Robert Schiffman for indispensable aid in assembling a punched card table of the Bessel function J_1. The calculations were performed with punched cards and equipment contributed by The Manhattan Life Insurance Company and The National Bureau of Economic Research.

REFERENCES

[1] K. J. Arnold, "Spherical probability distributions," Dissertation, Massachusetts Institute of Technology, 1941.

[2] C. E. P. Brooks and N. Carruthers, *Handbook of Statistical Methods in Meteorology*, London, H. M. Stationery Office, 1953.

[3] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[4] Sir Ronald Fisher, "Dispersion on a sphere," *Proc. Roy. Soc.* London. Ser. A, Vol. 217 (1953), pp. 295–305.

[5] J. S. Forrest, "Variations in thunderstorm severity in Great Britain," *Quart. J. Roy. Meteor. Soc.*, Vol. 76 (1950), pp. 277–286.

[6] E. J. Gumbel, J. A. Greenwood, and D. Durand, "The circular normal distribution: theory and tables," *J. Amer. Stat. Assn.*, Vol. 48 (1953), pp. 131–152.

[7] J. C. Kluyver, "A local probability problem," *Nederl. Akad. Wetensch., Proc.* Vol. 8 (1906), pp. 341–350.

[8] W. C. Krumbein, "Preferred orientation of pebbles in sedimentary deposits," *J. Geol.*, Vol. 47 (1939), pp. 673–706.

[9] P. Lévy, "Sur l'arithmétique des lois de probabilités enroulées," *C. R. Soc. Math. France*, 1938, pp. 32–34.

[10] P. Lévy, "L'addition des variables aléatoires définies sur une circonférence," *Bull. Soc. Math. France*, Vol. 67 (1939), pp. 1–41.

[11] J. M. Marcinkiewicz, "Sur les variables aléatoires enroulées," *C. R. Soc. Math. France*, 1938, pp. 34–36.

[12] R. v. Mises, "Über die 'Ganzzahligkeit' der Atomgewichte und verwandte Fragen," *Physikal. Zeitschr.*, Vol. 19 (1918), pp. 490–500.

[13] K. Pearson, "The problem of the random walk," *Nature*, Vol. 72 (1905), pp. 294, 302.

[14] K. Pearson, "A mathematical theory of random migration," *Drapers' Company Research Memoirs*, Biometric Series, No. 3, 1906.

[15] F. Perrin, "Étude mathématique du mouvement brownien de rotation," *Ann. Ecole Norm. Sup.*, Vol. 45 (1928), pp. 1–51.

[16] J. W. Strutt, Lord Rayleigh, "The problem of the random walk," *Nature*, Vol. 72 (1905), p. 318.

[17] J. W. Strutt, Lord Rayleigh, "On the problem of random vibrations and of random flights in one, two, and three dimensions," *Philos. Mag.* (6), Vol. 37 (1919), pp. 321–347.

[18] A. Wintner, "On the stable distribution laws," *Amer. J. Math.*, Vol. 55 (1933), pp. 335–339.

[19] A. Wintner, "On the shape of the angular case of Cauchy's distribution curves," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 589–593.

[20] F. Zernike, "Wahrscheinlichkeitsrechnung und mathematische Statistik," *Handbuch der Physik*, Berlin, Springer, Vol. 3 (1928), pp. 419–492.

ON THE EFFICIENCY OF EXPERIMENTAL DESIGNS[1]

BY SYLVAIN EHRENFELD

Columbia University

1. Introduction. Many models used in statistical investigations can be formulated in terms of least-square theory. The models that will be discussed can be stated as follows. Let $y_1, \cdots, y_N$ be N independently and normally distributed random variables with common variance σ^2. It is assumed that the expected value of y_α is given by

$$E(y_\alpha) = \beta_1 x_{\alpha 1} + \beta_2 x_{\alpha 2} + \cdots + \beta_p x_{\alpha p}, \qquad \alpha = 1, \cdots, N, \tag{1.1}$$

where quantities $x_{\alpha i}$ for $i = 1, \cdots, p; \alpha = 1, \cdots, N$ are known constants and $\beta_1, \cdots, \beta_p$ are unknown constants. The coefficients $\beta_1, \cdots, \beta_p$ are the population regression coefficients of y on $x_1, x_2, \cdots, x_p$ respectively. In matrix notation the above model can be expressed as

$$E(y) = X\beta \tag{1.2}$$

$$\beta = \begin{bmatrix} \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_p \end{bmatrix}, \quad X = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ x_{N1} & \cdots & x_{Np} \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_N \end{bmatrix},$$

The p column vectors in X will be denoted by $x_1, x_2, \cdots, x_p$ where $x_j' = (x_{1j}, x_{2j}, \cdots, x_{Nj})$. In some cases the experimenter has some amount of freedom in the choice of the p vectors x_i. The efficiency and sensitivity of the design may be very much affected by the choice of the design matrix X. The choice of this matrix is equivalent to that of p vectors in N-dimensional Euclidean space.

A simple illustration is furnished by the following example. Suppose y_α are independent random variables with equal variance σ^2, where $E(y_\alpha) = \beta_1(x_\alpha - \bar{x}) + \beta_2$. The x's are assumed to be fixed constants. Suppose, furthermore, that we have N pairs of observations $(x_1, y_1), \cdots, (x_N, y_N)$ and we want to estimate β_1. It is known that the variance of the least square estimate of β_1 is inversely proportional to $\sum_\alpha (x_\alpha - \bar{x})^2$. Hence, if we could choose values $x_1, \cdots, x_N$ in a domain T, we would choose them such that $\sum_\alpha (x_\alpha - \bar{x})$ is as large as possible.

In Section 2 we will prove a theorem about quadratic forms which, with the aid of other considerations, will motivate a criterion for the efficiency of a design matrix X. In Section 3 two theorems will be proved to aid in the applica-

Received July 3, 1953, revised May 7, 1954.

[1] Work partially sponsored by the Office of Naval Research.

tion of this criterion. In Section 4 applications will be given, and it will furthermore be shown that the Latin square is most efficient in a certain sense.

2. Measure of efficiency of design matrix.

2A. *A useful inequality.* Suppose we have a real $p \times p$ symmetric non-negative matrix S of rank $r \leqq p$. Let t be a column vector with p components, not all zero, such that the equation $S\rho = t$ has solutions for ρ. Then we have the following.

THEOREM 2.1. *If ρ is any solution of $S\rho = t$, then the following inequality holds*:

$$\frac{1}{\lambda_{\max}} \leqq \frac{\rho' S\rho}{t't} \leqq \frac{1}{\lambda_{\min}}, \tag{2.1}$$

where $\lambda_{\max}$ and $\lambda_{\min}$ are the maximum and the minimum of the nonzero characteristic roots of S.

PROOF. There exists an orthogonal matrix O such that $O'SO = \Omega = (\lambda_i\delta_{ij})$, and therefore $\rho'S\rho = \sum_1^r \lambda_i\rho_i^{*2}$. The $\lambda_1, \cdots, \lambda_r$ are the nonzero characteristic roots of S, and $\rho^* = O'\rho$. Since $S\rho = t$, we also have that $t't = \sum_1^r \lambda_i^2\rho_i^{*2}$ since $t't = \rho^{*\prime}O'S'OO'SO\rho^* = \rho^{*\prime}\Omega^2\rho^*$. Thus if we let $z_i = \sqrt{\lambda_i}\rho_i^*$, we have

$$\frac{\rho'S\rho}{t't} = \frac{\sum_i \lambda_i\, \rho_i^{*2}}{\sum_i \lambda_i^2\, \rho_i^{*2}} = \frac{\sum_i z_i^2}{\sum_i \lambda_i\, z_i^2}, \tag{2.2}$$

and since

$$\frac{1}{\lambda_{\max}} \leqq \frac{\sum_i z_i^2}{\sum_i \lambda_i\, z_i^2} \leqq \frac{1}{\lambda_{\min}}, \tag{2.3}$$

we get equation (2.1). When S is of full rank, equation (2.1) becomes

$$\frac{1}{\lambda_{\max}} \leqq \frac{t'S^{-1}\, t}{t't} \leqq \frac{1}{\lambda_{\min}}. \tag{2.4}$$

For the most part we will restrict ourselves to the case where S has full rank. In that case S has an inverse, and $\rho = S^{-1}t$. The case where S is not of full rank can be treated in a way very similar to the case of full rank. Some of the problems that arise in connection with the model which was considered in Section 1 will be discussed.

2B. *Estimation.* Suppose $S = X'X$ is of full rank, and we want to estimate θ, where $\theta = \sum_1^p t_j\beta_j = t'\beta$ and the t_j are given constants.

From least-square theory it is known that the best unbiased linear estimate of θ, in the sense of minimum variance, is $\hat{\theta} = \sum_j t_j\hat{\beta}_j$ where the $\hat{\beta}_j$ are the least-square estimates of β_j. Also, it is known that the variance of $\hat{\theta}$ is $\sigma^2t'S^{-1}t$. If t is such that $S\rho = t$ has solutions for ρ, we say that $\theta = t'\beta$ is estimable. Only estimable θ are here considered. When S is not of full rank, the variance of $\hat{\theta}$ is $\sigma^2\rho'S\rho$ where ρ is *any* solution of $S\rho = t$.

2C. *Power*. We will now derive an expression that will be used in Section 4. Suppose S and S^{-1} are partitioned into

$$\left(\begin{array}{c|c} a & c' \\ \hline c & d \end{array}\right) = S, \tag{2.5}$$

$$\left(\begin{array}{c|c} A & B \\ \hline B' & D \end{array}\right) = S^{-1}, \tag{2.6}$$

where a and A are $k \times k$ matrices. If we wanted to test $\beta_1 = \beta_2 = \cdots = \beta_k = 0$, then the usual F test has a power function depending monotonically on a parameter Φ where

$$\sigma^2\Phi = \bar{\beta}'A^{-1}\bar{\beta}, \qquad \bar{\beta}' = (\beta_1, \cdots, \beta_k). \tag{2.7}$$

Since $A^{-1} = a - c'd^{-1}c$, we have

$$\sigma^2\Phi = \bar{\beta}'a\bar{\beta} - (c\bar{\beta})'d^{-1}(c\bar{\beta}). \tag{2.8}$$

2D. *Confidence interval.* If $\sum_j t_j = 0$, then θ is called a contrast. A confidence interval for θ with confidence coefficient $1 - \alpha$ is

$$\hat{\theta} - k_\alpha\hat{\sigma}_{\hat{\theta}} \leqq \theta \leqq \hat{\theta} + k_\alpha\hat{\sigma}_{\hat{\theta}}, \tag{2.9}$$

where k_α is an appropriate constant. Here $\hat{\sigma}_{\hat{\theta}}^2$ is defined by

$$\hat{\sigma}_{\hat{\theta}}^2 = \hat{\sigma}^2 t'S^{-1}t, \tag{2.10}$$

where $\hat{\sigma}^2$ is the usual estimate of σ^2. If we let L equal the length of the confidence interval, we have

$$(E(L))^2 = M\sigma^2 t'S^{-1}t, \tag{2.11}$$

where M is independent of σ^2.

2E. *Application of inequality*. A simple application of Theorem 2.1 to some of the previous expressions gives

$$\frac{\sigma^2}{\lambda_{\max}} t't \leqq \operatorname{Var}(\hat{\theta}) \leqq \frac{\sigma^2}{\lambda_{\min}} t't, \tag{2.12}$$

$$\frac{M\sigma^2}{\lambda_{\max}} t't \leqq (E(L))^2 \leqq \frac{M\sigma^2}{\lambda_{\min}} t't. \tag{2.13}$$

If we wanted to test $\beta_1 = \cdots = \beta_p = 0$, we would get

$$\phi = \frac{\beta'S\beta}{\sigma^2}, \qquad \beta'\beta\frac{\lambda_{\max}}{\sigma^2} \geqq \Phi \geqq \frac{\lambda_{\min}}{\sigma^2}\beta'\beta. \tag{2.14}$$

It is to be noted that all these bounds can be attained. If (2.12), (2.13), and (2.14) are considered, it can be seen that it would be desirable to make $\lambda_{\min}$ as large as possible. With this in mind, we define the following as a criterion for the efficiency of a design.

DEFINITION. Denote by u the maximum value of $\lambda_{\min}$ for $x_{\alpha i}$ in T. The ratio of $\lambda_{\min}/u$ will be called the *efficiency* of the design. The design will be called *most efficient* if the efficiency of the design is equal to one.

3. Extension. In order to apply this criterion of efficiency, it will be helpful to have the following two theorems.

THEOREM 3.1. *If $x_i'x_i = c_i$ where $c_i \geqq 0$ are fixed constants, then $\lambda_{\min}$ will be a maximum when $x_i'x_j = 0$ for $i \neq j$.*

PROOF. Let c be the smallest of $c_1, \cdots, c_p$ and designate

$$S(\lambda) = S - \lambda I, \qquad \bar{S}(\lambda) = \begin{bmatrix} c_1 - \lambda & & 0 \\ & \cdot & \\ & \cdot & \\ & \cdot & \\ 0 & & c_p - \lambda \end{bmatrix},$$

and let $|S(\lambda)| = |S - \lambda I|$ be the determinant of $S(\lambda)$. We have that $|\bar{S}(\lambda)| = \prod_i(c_i - \lambda)$. Since $S(0)$ and $\bar{S}(0)$ are symmetric and non-negative, we also have $|S(0)| \geqq 0$ and $|\bar{S}(0)| \geqq 0$. Let us now consider the matrix $S(c)$. Suppose first that $S(c)$ is non-negative. We then have $0 \leqq |S(c)| \leqq |\bar{S}(c)| = 0$, and thus c would be a root of $S(0)$ and thus $\lambda_{\min} \leqq c$. Suppose, however, that $S(c)$ were not non-negative. Then there exists a root $\bar{\lambda} \leqq 0$ such that $|S(c) - \bar{\lambda}I| = 0$, or equivalently $|S(0) - cI - \bar{\lambda}I| = |S(0) - (c + \bar{\lambda})I| = 0$. Thus $c + \bar{\lambda}$ is a root of $S(0)$. Since $\bar{\lambda} \leqq 0$, we have $c + \bar{\lambda} \leqq c$ which gives the desired result.

When $x_i'x_j = 0$ for $i \neq j$, x_i is called orthogonal to x_j. When the p vectors are mutually orthogonal, then c, which is the upper bound of $\lambda_{\min}$ can be attained. In some situations one can maintain the condition of orthogonality and increase c at the same time. This theorem is usually useful only if the experimenter has some freedom of choice in all vectors $x_1, x_2, \cdots, x_p$.

In many situations, however, the vectors $x_1, \cdots, x_r$ are fixed, and there is only freedom of choice of $x_{r+1}, \cdots, x_p$. In those cases the following theorem is sometimes useful. Its proof is very similar to the one used for Theorem 3.1.

Suppose $x_1, \cdots, x_r$ are fixed vectors. Then S can be represented as a partitioned matrix as follows:

$$S = S(b, D) = \left(\begin{array}{c|c} A & b \\ \hline b' & D \end{array}\right), \tag{3.1}$$

where $S(b, D)$ indicates that S depends on b and D. Also we have

$$\begin{aligned} A &= (x_i'x_j), \qquad & i, j &= 1, \cdots, r, \\ D &= (x_i'x_j), \qquad & i, j &= r + 1, \cdots, p. \end{aligned} \tag{3.2}$$

$S(b, D)$, A, and D are symmetric and non-negative matrices. Let the characteristic roots of A be $0 \leqq \lambda_1 \leqq \cdots \leqq \lambda_r$, and those of D be $0 \leqq \theta_1 \leqq \theta_2 \leqq \cdots \leqq \theta_{p-r}$. The λ's are fixed, but the θ's depend on the choice of $x_{r+1}, \cdots, x_p$.

THEOREM 3.2. *If ψ is the smallest characteristic root of $S = S(b, D)$, then*

$$\psi \leqq \lambda_1 \tag{3.3}$$

when $b = 0$ and $\lambda_1 \leqq \theta_1$, then $\psi = \lambda_1$.

PROOF. There exist orthogonal matrices T, Q such that

$$T'AT = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_r \end{bmatrix}, \quad Q'DQ = \begin{bmatrix} \theta_1 & & 0 \\ & \ddots & \\ 0 & & \theta_{p-r} \end{bmatrix}. \tag{3.4}$$

If we let

$$P = \begin{pmatrix} T & 0 \\ 0 & Q \end{pmatrix} \tag{3.5}$$

then $P'SP$ has the same characteristic roots as S, since P is an orthogonal matrix. We also get

$$P'SP = \left[\begin{array}{ccc|ccc} \lambda_1 & & 0 & & & \\ & \ddots & & & b^* & \\ 0 & & \lambda_r & & & \\ \hline & & & \theta_1 & & 0 \\ & b^{*\prime} & & & \ddots & \\ & & & 0 & & \theta_{p-r} \end{array}\right], \qquad b^* = T'bQ. \tag{3.6}$$

Since b^* is zero if and only if b is, we have by use of Theorem 3.1 that $\psi \leqq \min (\lambda_1 \cdots, \lambda_r, \theta_1, \cdots, \theta_{p-r}) \leqq \lambda_1$ which proves the theorem.

In terms of the vectors $x_{r+1}, \cdots, x_p$ Theorem 3.2 means that to maximize $\lambda_{\min}$ we must choose $x_i'x_j = 0$ for $i = 1, \cdots, r$; $j = r + 1, \cdots, p$. In order to attain the upper bound λ_1, we must also make $\lambda_1 \leqq \theta_1$.

To do this we might use Theorem 3.1, and make $x_t'x_s = 0$ for $t, s = r + 1, \cdots, p$ and $x_t'x_t \geqq \lambda_1$ for $t = r + 1, \cdots, p$ if possible.

4. Applications.

4A. *Hotelling's weighing problem.* Consider the weighing of N linear combination of p objects on a chemical balance. This gives rise to the following equations:

$$E(y_\alpha) = x_{\alpha 1}\omega_1 + \cdots + x_{\alpha p}\omega_p, \qquad \alpha = 1, \cdots, N, \tag{4.1}$$

or

$$E(y) = X\omega, \quad X = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{N1} & \cdots & x_{Np} \end{bmatrix} \tag{4.2}$$

where ω_i is the weight of the ith object.

$$x_{\alpha i} = \begin{cases} +1 & \text{if } i\text{th object is placed in left pan at } \alpha\text{th weighing,} \\ -1 & \text{if it is placed in right pan,} \\ 0 & \text{if it is placed in neither.} \end{cases} \tag{4.3}$$

Thus the vectors $x_1, x_2, \cdots, x_p$ are vectors whose components are $+1$, -1, or 0. From Theorem 3.1, it follows that the most efficient design for this problem would be gotten by making $x_i'x_j = 0$ for $i \neq j$ and to make $x_i'x_i$ as large as possible. This can be done by picking $x_{\alpha i} = +1$ or $x_{\alpha i} = -1$, then $x_i'x_i = N$. The vectors $x_1, \cdots, x_p$ cannot be chosen as indicated for all N. The problem of when and how this can be done has been thoroughly discussed in the literature [3].

4B. *Power*. From equation (2.8) it can be seen that $\min_{\beta'\beta\leq 1} \sigma^2\Phi$ would be a maximum if $c = 0$ and the minimum root of a is large. The condition $c = 0$ is equivalent to making $x_i'x_j = 0$ for $i = 1, \cdots, k; j = k + 1, \cdots, p$, and from Theorem 3.1 the $\lambda_{\min}$ of a can be made large by making $x_i'x_i$ for $i = 1, \cdots, k$ as large as possible and $x_s'x_t = 0$ for $s \neq t$; $s, t = 1, \cdots, k$. Any linear hypothesis can be put into canonical form so that it becomes $\beta_1 = \beta_2 = \cdots = \beta_k = 0$.

4C. *Analysis of covariance*. The model usually used for a two-way layout can be stated as follows.

$$E(y_{ij}) = u + Q_i + T_j + \beta(x_{ij} - \bar{x}), \qquad i = 1, \cdots, r; j = 1, \cdots, s, \tag{4.4}$$

where the Q's and T_j's are row and column affects respectively. Associate with each pair (i, j) an integer α such that $\alpha(i, j) = (i - 1)s + j$ for $\alpha = 1, \cdots, rs$. Then (4.4) can be written as

$$E(y_\alpha) = u + \textstyle\sum_i t_{i\alpha}Q_i + \sum_j v_{j\alpha}T_j + \beta(x_\alpha - \bar{x}) \tag{4.5}$$

where

$$\begin{aligned} t_{i\alpha} &= +1 \text{ if } \alpha\text{th plot is in } i\text{th row and 0 otherwise,} \\ v_{j\alpha} &= +1 \text{ if } \alpha\text{th plot is in } j\text{th column and 0 otherwise.} \end{aligned} \tag{4.6}$$

In matrix notation the above becomes

$$E(y) = X\,\theta = (1 \;\; t_{1\alpha} \;\; t_{2\alpha} \cdots t_{r\alpha} \;\; v_{1\alpha} \cdots v_{s\alpha} \;\; (x_\alpha - \bar{x})) \begin{bmatrix} u \\ Q_1 \\ \vdots \\ Q_r \\ T_1 \\ \vdots \\ T_s \\ \beta \end{bmatrix}. \tag{4.7}$$

It can be seen that $x_1, \cdots, x_{r+1}, x_{r+2}, \cdots, x_{r+s+1}$ are fixed vectors in the model. The only choice available to the experimenter is in the last vector x_{r+s+2}.

An application of Theorem 3.2 gives us the following advice for getting a good design. Make $x'_{r+s+2}x_{r+s+2}$ as large as possible, where

$$x'_{r+s+2}x_{r+s+2} = \sum_\alpha (x_\alpha - \bar{x})^2 = \sum_{i,j} (x_{ij} - \bar{x})^2. \tag{4.8}$$

Furthermore, it is desirable to make $x'_{r+s+2}x_j = 0$ for $j = 1, \cdots, r + s + 1$ simultaneously, that is;

$$\sum_\alpha t_{i\alpha}(x_\alpha - \bar{x}) = \sum_j (x_{ij} - \bar{x}) = 0, \tag{4.9}$$

$$\sum_\alpha v_{j\alpha}(x_\alpha - \bar{x}) = \sum_i (x_{ij} - \bar{x}) = 0. \tag{4.10}$$

Equations (4.9) and (4.10) thus give conditions for an efficient design for this model.

4D. *Latin square.* Suppose we wish to find out by experimentation whether there is any significant difference among yields of m different varieties $v_1, \cdots, v_m$. An experimental area is divided into m^2 plots lying in m rows and m columns and each plot is assigned to one of the varieties $v_1, \cdots, v_m$. Denote by y_{ijk} the yield of variety v_k on plot in ith row and jth column. It is assumed that the y_{ijk} are independently and normally distributed with common variance σ^2.

$$E(y_{ijk}) = k_i + d_j + \rho_k \tag{4.11}$$

where k_i, d_j, ρ_k are the row, column, and variety effects respectively. The quantities σ^2, k_i, d_j, ρ_k are unknown parameters. The hypothesis to be tested is that variety has no effect on yield, i.e.,

$$\rho_1 = \rho_2 = \cdots = \rho_k . \tag{4.12}$$

Wald proved that the Latin square was most efficient in a sense which he defined in [1]. We will prove a similar result in this section.

THEOREM 4.1. *The Latin square is most efficient in the sense of our definition.*

PROOF. As was done in (4B) we let $\alpha(i, j) = (i - 1)m + j$ for $i, j = 1, \cdots, m$. Let $t_{i\alpha}$, $u_{j\alpha}$, $z_{k\alpha}$ for $i, j, k = 1, \cdots, m$; $\alpha = 1, \cdots, m^2$ be defined as follows.

$$\begin{aligned} t_{i\alpha} &= +1 \text{ if } \alpha\text{th plot in } i\text{th row and 0 otherwise,} \\ u_{j\alpha} &= +1 \text{ if } \alpha\text{th plot in } j\text{th column and 0 otherwise,} \\ z_{k\alpha} &= +1 \text{ if } k\text{th variety is assigned to } \alpha\text{th plot.} \end{aligned} \tag{4.13}$$

Then (4.11) can be rewritten as follows.

$$E(y_\alpha) = \sum_{i=1}^m k_1 t_{i\alpha} + \sum_{j=1}^m d_j u_{j\alpha} + \sum_{k=1}^m \rho_k z_{k\alpha} . \tag{4.14}$$

We want to express the hypothesis in (4.12) in canonical form. In order to do this denote the arithmetic means

$$\bar{t}_1 = \frac{1}{m^2} \sum_\alpha t_{i\alpha} ; \qquad \bar{u}_i = \frac{1}{m^2} \sum_\alpha u_{i\alpha} ; \qquad \bar{z}_i = \frac{1}{m^2} \sum_\alpha z_{i\alpha} \tag{4.15}$$

and then let $t'_{i\alpha} = t_{i\alpha} - \bar{t}_i$, $u'_{i\alpha} = u_{i\alpha} - \bar{u}_i$, $z'_{i\alpha} = z_{i\alpha} - \bar{z}_i$ and $k'_i = k_i - k_m$, $d'_i = d_i - d_m$, $\rho'_i = \rho_i - \rho_m$ for $i = 1, \cdots, m - 1$, and let $\omega_\alpha = 1$ for $\alpha = 1, \cdots, m^2$. Then (4.14) can be written as

$$E(y_\alpha) = \xi\,\omega_\alpha + \sum_{i=1}^{m-1} k'_i t'_{i\alpha} + \sum_{i=1}^{m-1} d'_i u'_{i\alpha} + \sum_{i=1}^{m-1} \rho'_i z'_{i\alpha}, \tag{4.16}$$

where $\xi = \sum_1^{m-1} k'_i \bar{t}_i + \sum_1^{m-1} d'_i \bar{u}_i + \sum_1^{m-1} \rho'_i \bar{z}_i + k_m + d_m + \rho_m$, and the hypothesis in canonical form is

$$\rho'_1 = \rho'_2 = \cdots = \rho'_{m-1} = 0. \tag{4.17}$$

In matrix notation the design matrix X becomes

$$X = (1 \quad t'_{1\alpha} \quad t'_{2\alpha} \cdots t'_{m-1,\alpha} \quad u'_{1\alpha} \cdots u'_{m-1,\alpha} \quad z'_{1\alpha} \cdots z^{1}_{m-1,\alpha}). \tag{4.18}$$

The experimenter has freedom only in the choice of the $z'_{i\alpha}$, $i = 1, \cdots, m-1$; $\alpha = 1, \cdots, m^2$. The $z'_{i\alpha}$ depend on the way the varieties $v_1, \cdots, v_m$ are assigned to the m^2 plots. In the Latin square arrangement each variety appears exactly once in each row and exactly once in each column. The S matrix for the above model is

$$S = \left[\begin{array}{c|c|c|c} m^2 & 0 & 0 & \\ \hline 0 & \begin{matrix} m-1 & & -1 \\ & \ddots & \\ -1 & & m-1 \end{matrix} & 0 & b \\ \hline 0 & 0 & \begin{matrix} m-1 & & -1 \\ & \ddots & \\ -1 & & m-1 \end{matrix} & \\ \hline & b' & & D \end{array}\right] \tag{4.19}$$

and we let $D = (z''_i z'_j)$ and

$$A = \left[\begin{array}{c|c|c} m^2 & 0 & 0 \\ \hline 0 & \begin{matrix} m-1 & & -1 \\ & \ddots & \\ -1 & & m-1 \end{matrix} & 0 \\ \hline 0 & 0 & \begin{matrix} m-1 & & -1 \\ & \ddots & \\ -1 & & m-1 \end{matrix} \end{array}\right]. \tag{4.20}$$

It is known that for the Latin square

$$\textstyle\sum_\alpha z'_{k\alpha} u'_{j\alpha} = \sum_\alpha z'_{k\alpha} t'_{i\alpha} = \sum_\alpha z'_{k\alpha}\omega_\alpha = 0, \qquad i, j, k = 1, \cdots, m - 1, \tag{4.21}$$

which is exactly the condition of orthogonality needed in Theorem 3.2, namely $b = 0$. For a Latin square the D matrix becomes

$$D = \begin{bmatrix} m-1 & & -1 \\ & \cdot & \\ & \cdot & \\ & \cdot & \\ -1 & & m-1 \end{bmatrix}, \tag{4.22}$$

and if we use the fact that

$$|Q| = \begin{vmatrix} a-1 & & -1 \\ & \cdot & \\ & \cdot & \\ & \cdot & \\ -1 & & a-1 \end{vmatrix} = a^{k-1}[a-k],$$

where Q is a $k \times k$ matrix, it is thus seen that $\lambda_1 \leqq \theta_1$, which proves the theorem.

Acknowledgement. I would like to thank Professor Henry Scheffé for many helpful suggestions.

REFERENCES

[1] A. Wald, "On efficient design of statistical investigations," *Ann. Math. Stat.*, Vol. 14 (1943), pp. 134–140.

[2] O. Schreier and E. Sperner, *Introduction to Modern Algebra and Matrix Theory*, Chelsea Publishing Company, New York, 1951.

[3] A. M. Mood, "On Hotelling's weighing problem," *Ann. Math. Stat.*, Vol. 17 (1946), pp. 432–447.

[4] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, Cambridge, 1934, Section 2.13.

[5] C. R. Rao, "On the linear combination of observations and the general theory of least squares," *Sankhyā*, Vol. 7 (1946), pp. 237–256.

[6] F. N. David and J. Neyman, "Extension of the Markoff theorem of least squares," *Stat. Res. Memoirs*, Vol. 2 (1938), pp. 105–116.

ON THE APPROXIMATION OF A DISTRIBUTION FUNCTION BY AN EMPIRIC DISTRIBUTION[1]

By Jerome Blackman

Syracuse University

1. Summary. Let $x_1, \cdots, x_n$ be independent chance variables with the common distribution function $F(x)$ and the empiric distribution function $F^*(x)$. Let a_n be the value of a which minimizes (1) below. In this paper the asymptotic distribution of $\sqrt{n}\, a_n$ is obtained, subject to certain restrictions on $F(x)$.

2. Introduction. Let $x_1, x_2, \cdots, x_n$ be n independent random variables having the common distribution $F(x)$. Suppose $b_1, b_2, \cdots, b_n$ are the n random variables in ascending order of values and $F^*(x)$ is the empiric distribution function, continuous on the right, with jumps of magnitude $1/n$ at the points $b_1, b_2, \cdots, b_n$. Define the function $H(a)$ by

$$H(a) = \int_{-\infty}^{\infty} |F(x-a) - F^*(x)|^2 \, dF(x-a). \tag{1}$$

Thus $H(a)$ is non-negative for all a and since it is a Borel-measurable function of the random variables $\{x_i\}$, it is also a random variable. The value of a which minimizes $H(a)$ will also be a random variable. If the minimizing value of a is a_n, we shall be concerned with the limiting distribution of a_n as $n \to \infty$. Our main result is the following.

Theorem 2. *If the first three derivatives of $F(x)$ are continuous and bounded, then*

$$\lim_{n\to\infty} P(\sqrt{n}\, a_n < x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\mu x/\nu} e^{-y^2/2}\, dy,$$

where

$$\nu^2 = \int_0^1 \left| \int_0^y F'(F^{-1}(t))\, dt - \int_0^y \frac{dx}{x^2} \int_0^x tF'(F^{-1}(t))\, dt \right|^2 dy,$$

$$\mu = \int_{-\infty}^{\infty} |F'(x)|^2 \, dF(x), \qquad F^{-1}(t) = \sup_x \{x \mid F(x) = t\}.$$

We shall henceforth assume the conditions of Theorem 2 are satisfied. In what follows repeated use will be made of an important result due to Kolmogoroff [2].

Theorem. *Suppose that $F(x)$ is continuous, and define the random variable D_n by $D_n = \text{l.u.b.}_x |F(x) - F^*(x)|$.*

Received June 17, 1954.

[1] Work on this paper was done under sponsorship of the Office of Naval Research. The author would like to thank Professor J. Wolfowitz for many useful suggestions and criticisms in the preparation of this paper.

Then for every fixed $z \geq 0$, as $n \to \infty$, $P(D_n \leq zn^{-1/2}) \to L(z)$, where $L(z)$ is the cumulative distribution function which for $z > 0$ is given by either of the equivalent relations

$$L(z) = 1 - 2\sum_{\nu=1}^{\infty} (-1)^{\nu-1} \exp(-2\nu^2 z^2) = \frac{\sqrt{2\pi}}{z} \sum_{\nu=1}^{\infty} \exp\left\{-\frac{(2\nu-1)^2 \pi^2}{8z^2}\right\}.$$

For $z \leq 0$ we have, of course, $L(z) = 0$.

The Kolmogoroff theorem implies that $F^*(x) \to F(x)$ uniformly in probability and that therefore $a_n \to 0$ in probability. Expanding the right side of (1),

$$\begin{aligned} H(a) &= \int_{-\infty}^{\infty} \{F(x-a) - F^*(x)\}^2 \, dF(x-a) \\ &= \int_{-\infty}^{\infty} F^2(x-a)\, dF(x-a) - 2\int_{-\infty}^{\infty} F(x-a)\, F^*(x)\, dF(x-a) \\ &\qquad + \int_{-\infty}^{\infty} F^{*2}(x)\, dF(x-a) \\ &= \tfrac{1}{3} + \int_{-\infty}^{\infty} F^2(x-a)\, dF^*(x) - \int_{-\infty}^{\infty} F(x-a)\, dF^{*2}(x). \end{aligned} \tag{2}$$

Therefore

$$H'(a) = \left\{-2\int_{-\infty}^{\infty} F(x-a)\, F'(x-a)\, dF^* + \int_{-\infty}^{\infty} F'(x-a)\, dF^{*2}(x)\right\}. \tag{3}$$

Each a_n must satisfy the equation $H'(a) = 0$. It will be seen below that all solutions of $H'(a) = 0$ which converge in probability to zero as $n \to \infty$ will have the same limiting distribution. Putting $H(a_n) = 0$ we obtain

$$2n^{-1} \sum F(b_i - a_n) F'(b_i - a_n) = n^{-2} \sum \{i^2 - (i-1)^2\} F'(b_i - a_n),$$

where summation is from 1 to n, or

$$\sum \{2nF(b_i - a_n) - 2i + 1\} F'(b_i - a_n) = 0. \tag{4}$$

Since $F(x)$ has a continuous third derivative,

$$\begin{cases} F(b_i - a_n) = F(b_i) - F'(b_i)a_n + \frac{1}{2}F''(b_i - \theta a_n)\, a_n^2, & 0 \leq \theta \leq 1; \\ F'(b_i - a_n) = F'(b_i) - F''(b_i)a_n + \frac{1}{2}F'''(b_i - \psi a_n)a_n^2, & 0 \leq \psi \leq 1. \end{cases} \tag{5}$$

Placing (5) in (4) and dividing by n^2 results in

$$\begin{aligned} 0 = \{&-2n^{-1}\textstyle\sum F(b_i)F'(b_i) - n^{-2}\sum(-2i+1)F'(b_i)\} \\ &+ \{2n^{-1}\textstyle\sum F(b_i)F''(b_i) + 2n^{-1}\sum F'^2(b_i) + n^{-2}\sum(-2_i + 1)F''(b_i)\}a_n \\ &+ T_n(a_n)a_n^2 \end{aligned} \tag{6}$$

where $P\{|T(a_n)| \leq C\} \geq 1 - \epsilon$ for some C and for $n > N(\epsilon)$ by the assump-

tion that the derivatives are bounded and that $a_n \to 0$ in probability. Equation (6) is of the form

$$A_n + B_n a_n + T_n(a_n) a_n^2 = 0. \tag{7}$$

Let a_n' be the solution of

$$A_n + B_n a_n' = 0. \tag{8}$$

Subtracting (8) from (7) gives that, with probability $\geqq 1 - \epsilon$ for $n > N(\epsilon)$,

$$|a_n - a_n'| = \frac{|T_n(a_n)|\, a_n^2}{|B_n|} \leqq \frac{C a_n^2}{|B_n|}. \tag{9}$$

It will be shown below that $B_n \to \mu \neq 0$ in probability, so that

$$|a_n - a_n'| \,/\, |a_n| \to 0 \text{ in probability.}$$

It is therefore necessary only to find the limiting distribution of a_n' where

$$\begin{aligned} a_n' = \frac{-A_n}{B_n} &= \frac{n^{-1} \sum [F(b_i) - i/n + 1/2n]\, F'(b_i)}{n^{-1} \sum [F(b_i) - i/n + 1/2n]\, F''(b_i) + n^{-1} \sum F'^2(b_i)} \\ &= \frac{n^{-1} \sum [F(b_i) - i/n] F'(b_i) + \frac{1}{2} n^{-2} \sum F'(b_i)}{n^{-1} \sum [F(b_i) - i/n]\, F''(b_i) + n^{-1} \sum F'^2(b_i) + \frac{1}{2} n^{-2} \sum F''(b_i)}. \end{aligned} \tag{10}$$

3. Lemmas. Three lemmas are useful in the proof of the main result, Theorem 2.

LEMMA 1. *For every* $\epsilon > 0$

$$\lim_{n\to\infty} P(\,|n^{-1/2}\, F'(b_i)[F(b_i) - i/n] - n^{-1/2} \sum F'(F^{-1}(i/n))[F(b_i) - i/n]| > \epsilon) = 0.$$

PROOF. Since

$$\begin{aligned} &n^{1/2}\{n^{-1} \sum |F'(b_i) - F'(F^{-1}(i/n))|\, |F(b_i) - i/n|\,\} \\ &\qquad \leqq \text{l.u.b.}_i\, |F'(b_i) - F'(F^{-1}(i/n))|\ \text{l.u.b.}_i\, |F(b_i) - i/n|\, n^{1/2}, \end{aligned}$$

we need, by the Kolmogoroff theorem, show only that

$$\text{l.u.b.}_i\, |F'(b_i) - F'(F^{-1}(i/n))| \to 0 \text{ in probability.}$$

Suppose $\epsilon > 0$ and $\eta > 0$ given. Let A be the linear set such that $F'(x) > \eta/4$ for $x \,\varepsilon\, A$ and let A' be the complement of A. Let $D = (b_i, F^{-1}(i/n))$ be the open interval with end points b_i and $F^{-1}(i/n)$, and $M = \text{l.u.b.}_x\, |F''(x)|$. By the Kolmogoroff Theorem there is an N such that for $n > N$

$$P(\text{l.u.b.}_i\, |F(b_i) - i/n| \geqq \eta^2/16M) < \epsilon. \tag{11}$$

$$\text{l.u.b.}_i\, |F(b_i) - i/n| \geqq \left| \int_{b_i}^{F^{-1}(i/n)} F'(x)\, dx \right| \geqq \left| \int_{D \cap A} F'(x)\, dx \right| \geqq \tfrac{1}{4}\eta \operatorname{meas}[D \cap A].$$

On the set of sample points for which

$$\text{l.u.b.}_i\, |F(b_i) - i/n| < \eta^2/16M$$

we therefore have

$$\text{meas } [D \cap A] < \eta/4M \quad \text{for all} \quad i \leq n.$$

Either meas $(D) < \eta/4M$, in which case $|F'(b_i) - F'(F^{-1}(i/n))| \leq (\eta/4M)M < \eta$, or there are points γ_i and δ_i for all $i \leq n$ such that $\gamma_i, \delta_i \varepsilon A'$,

$$|\gamma_i - b_i| < \eta/4M, \qquad |\delta_i - F^{-1}(i/n)| < \eta/4M.$$

Therefore, in the second case

$$\begin{aligned}
&|F'(b_i) - F'(F^{-1}(i/n))| \\
&\qquad \leq |F'(b_i) - F'(\gamma_i)| + |F'(\gamma_i) - F'(\delta_i)| + |F'(\delta_i) - F'(F^{-1}(i/n))| \\
&\qquad \leq M\,|b_i - \gamma_i| + 2\eta/4 + M\,|\delta_i - F^{-1}(i/n)| \\
&\qquad \leq \eta/4 + \eta/2 + \eta/4 = \eta.
\end{aligned}$$

Therefore this inequality is true in any case except for the set of equations (11). Combining our results we have for all $n > N$

$$P(\text{l.u.b.}_i\, |F'(b_i) - F'(F^{-1}(i/n))| > \eta) < \epsilon,$$

which proves the lemma.

LEMMA 2. *Let* $H_0(x) = F'(F^{-1}(x))$ for $0 \leq x \leq 1$ *and*

$$H_n(x) = \frac{1}{x}\int_0^x H_{n-1}(t)\, dt. \tag{12}$$

Then both

(a) $\int_0^1 x^{-1} H_n(x)\, dx < \infty$,

(b) $\sum_{k=1}^{\infty} (-1)^{k+1} H_k(x) = \frac{1}{x^2}\int_0^x t\, H_0(t)\, dt$, *uniformly in* x.

PROOF. Let $M = \text{l.u.b.}_x\, |F''(x)|$. Consider the curve $y = F'(x)$ at the point $(x, F'(x))$. If from this point a line of slope M is drawn to the x-axis, the line must be completely on or below the curve $y = F'(x)$ in $(-\infty, x)$, by the mean-value theorem. Since $F(x) = \int_{-\infty}^{x} F'(x)\, dx$, it follows that

$$F(x) \geq \tfrac{1}{2}F'^2(x)/M. \tag{13}$$

This implies that $\lim_{x\to 0+} H_0(x) = 0$ and by an obvious induction that $\lim_{x\to 0+} H_n(x) = 0$.

Differentiating (12) gives $H_n'(x) = -x^{-1}H_n(x) + x^{-1}H_{n-1}(x)$, or

$$\int_0^y x^{-1} H_n(x)\, dx = \int_0^y x^{-1} H_{n-1}(x)\, dx - H_n(y). \tag{14}$$

Therefore the truth of (a) for $n - 1$ implies it for n and we need show $\int_0^1 x^{-1}H_0(x)\, dx = \int_{-\infty}^{\infty} (F'(x)/F(x))\, dF < \infty$. Define the set E by

$$E = \{x \mid F'(x)/F(x) > F^{-1/2}(x)\}.$$

If E' is the complementary set,

$$(15)\qquad \int_{E'} (F'(x)/F(x))\, dF \leqq \int_{E'} F^{-1/2}(x) dF \leqq \int_0^1 x^{-1/2}\, dx = 2.$$

The set E consists of a union of open intervals, by the continuity of F and F'. Let $[a, b]$ be one such interval. On $[a, b]$

$$F'(x \geqq F^{1/2}(x), \qquad \int_a^b F'(x)/F^{1/2}(x)\, dx \geqq \int_a^b dx$$

or $2[F^{1/2}(b) - F^{1/2}(a)] \geqq b - a$. This implies that the measure of E is $\leqq 2$. Using the inequality (13)

$$(16)\qquad \int_E (F'(x)/F(x))\, dF = \int_E (F'^2(x)/F(x))\, dx \leqq 2M \int_E dx \leqq 4M.$$

This completes the proof of (a). Turning to (b) we note that by (14) $\int_0^y x^{-1} H_n(x)\, dx$ is a decreasing positive function of n for each y and therefore has a limit as $n \to \infty$. Taking this limit in (14) shows that $\lim_{n\to\infty} H_n(y) = 0$ for all y. Writing

$$\int_0^1 x^{-1} H_n(x)\, dx = \int_0^\eta x^{-1} H_n(x)\, dx + \int_\eta^1 x^{-1} H_n(x)\, dx$$
$$\leqq \int_0^\eta x^{-1} H_0(x)\, dx + \int_\eta^1 x^{-1} H_n(x)\, dx$$

we see that η can be chosen so as to make the first of these arbitrarily small, and that the second then goes to zero by bounded convergence. Therefore $\lim_{n\to\infty} \int_0^1 x^{-1} H_n(x)\, dx = 0$. From (14) we obtain

$$\textstyle\sum_1^n H_k(x) = \int_0^x t^{-1} H_0(t)\, dt - \int_0^x t^{-1} H_n(t)\, dt.$$

Since $|\int_0^x t^{-1} H_n(t)\, dt| \leqq \int_0^1 t^{-1} H_n(t)\, dt \to 0$ it follows that $\sum_1^\infty H_k(x)$ converges uniformly to $\int_0^x t^{-1} H_0(t)\, dt$. Let $J_n(x) = \sum_1^n (-1)^{k+1} H_k(x)$. Then $J_n(x) \to J(x)$ uniformly as $n \to \infty$, and

$$\begin{aligned} J_n' &= \textstyle\sum_1^n (-1)^{k+1} H_k'(x) \\ &= x^{-1}[H_0 - H_1 - H_1 + H_2 \cdots (-1)^{n+1} H_{n-1} + (-1)^n H_n] \\ &= x^{-1} H_0 - 2x^{-1} J_{n-1} + x^{-1}(-1)^k H_n . \end{aligned}$$

The right side converges uniformly for $\epsilon \leqq x \leqq 1$ so that for $0 < x \leqq 1$,

$$J' + 2x^{-1} J = x^{-1} H_0(x),\ x^2 J' + 2xJ = xH_0(x),\ d/dx(x^2 J) = xH_0(x),$$

or $J(x) = x^{-2} \int_0^x tH_0(t)\, dt$, proving (b) and thus Lemma 2.

LEMMA 3. *For every* $\epsilon > 0$ *and* $s = 0, 1, 2, \cdots$

$$\lim_{n\to\infty} P\left[\left|\frac{1}{\sqrt{n}}\sum_1^n H_s\left(\frac{i}{n}\right)\left\{F(b_i) - \frac{i}{n}\right\} - \sqrt{n}\sum_1^n \int_0^{i/n} H_s(x)\,dx\right.\right.$$

$$\left.\left.\left\{\frac{F(b_i) - F(b_i+1)}{F(b_{i+1})}\frac{i+1}{n} + \frac{1}{n}\right\} + \frac{1}{\sqrt{n}}\sum\nolimits_1^n H_{s+1}\left(\frac{i}{n}\right)\left\{F(b_i) - \frac{i}{n}\right\}\right| > \epsilon\right] = 0.$$

Proof. In what follows we take $F(b_{n+1}) = 1$ whenever it occurs. Letting $G_k = \sum_{i=1}^k H_s(i/n)$ and using the Abel transformation,

$$\frac{1}{\sqrt{n}}\sum_1^n H_s\left(\frac{i}{n}\right)\left\{F(b_i) - \frac{i}{n}\right\}$$

$$= \frac{1}{\sqrt{n}}\sum_1^{n-1} G_i\left\{F(b_i) - F(b_{i+1}) + \frac{1}{n}\right\} + \frac{1}{\sqrt{n}} G_n\{F(b_n) - 1\}$$

$$= \sqrt{n}\sum_1^{n-1}\int_0^{i/n} H_s(x)\,dx\left\{F(b_i) - F(b_{i+1}) + \frac{1}{n}\right\}$$

$$+ \sqrt{n}\sum_1^{n-1}\left\{\frac{1}{n}G_i - \int_0^{i/n} H_s(x)\,dx\right\}\left\{F(b_i) - F(b_{i+1}) + \frac{1}{n}\right\}$$

$$+ \frac{1}{\sqrt{n}} G_n\{F(b_n) - 1\}$$

$$= \sqrt{n}\sum_1^{n-1}\int_0^{i/n} H_s(x)\,dx\left\{\frac{F(b_i) - F(b_{i+1})}{F(b_{i+1})}\,\frac{i+1}{n} + \frac{1}{n}\right\}$$

$$+ \sqrt{n}\sum_1^{n-1}\left\{\frac{1}{n}G_i - \int_0^{i/n} H_s(x)\,dx\right\}\left\{F(b_i) - F(b_{i+1}) + \frac{1}{n}\right\}$$

$$+ \sqrt{n}\sum_1^{n-1}\int_0^{i/n} H_s(x)\,dx\left\{\frac{F(b_i) - F(b_{i+1})}{F(b_{i+1})}\right\}\left\{F(b_{i+1}) - \frac{i+1}{n}\right\}$$

$$+ \frac{1}{\sqrt{n}} G_n\{F(b_n) - 1\} \tag{16}$$

$$= \sqrt{n}\sum_1^{n}\int_0^{i/n} H_s(x)\,dx\left\{\frac{F(b_i) - F(b_{i+1})}{F(b_{i+1})}\,\frac{i+1}{n} + \frac{1}{n}\right\}$$

$$+ \sqrt{n}\sum_1^{n-1}\left\{\frac{1}{n}G_i - \int_0^{i/n} H_s(x)\,dx\right\}\left\{F(b_i) - F(b_{i+1}) + \frac{1}{n}\right\}$$

$$+ \sqrt{n}\sum_1^{n}\int_0^{i/n} H_s(x)\,dx\left\{\frac{F(b_i) - F(b_{i+1})}{(i+1)/n}\right\}\left\{F(b_{i+1}) - \frac{i+1}{n}\right\}$$

$$+ \sqrt{n}\sum_1^{n}\int_0^{i/n} H_s(x)\,dx\,\{F(b_i) - F(b_{i+1})\}$$

$$\left\{\frac{1}{F(b_{i+1})} - \frac{n}{i+1}\right\}\left\{F(b_{i+1}) - \frac{i+1}{n}\right\}$$

$$- \sqrt{n} \int_0^1 H_s(x)\, dx\, \{F(b_n) - 1\} \left\{1 - \frac{n+1}{n}\right\}$$

$$- \sqrt{n} \int_0^1 H_s(x)\, dx \left\{(F(b_n) - 1)\frac{n+1}{n} + \frac{1}{n}\right\} + \frac{1}{\sqrt{n}} \{F(b_n) - 1\}\, G_n$$

$$= \sum_1^7 A_i\,.$$

We first must show that A_2, A_4, A_5, A_6, and A_7 converge to zero in probability. That this is so for A_5, A_6, and A_7 is a simple consequence of the fact that since $F(b_n)$ represents the maximum of n uniformly distributed independent random variables, the distribution function of $1 - F(b_n)$ is $1 - (1 - x)^n$ for $0 \leqq x \leqq 1$, from which it follows that $n^{1-\eta}[1 - F(b_n)]$ converges to zero in probability for $\eta > 0$. Of course, $n^{-1}G_n$ is bounded.

Using the Abel transformation

$$A_2 = \sqrt{n} \sum_1^{n-2} \left\{ \int_{i/n}^{(i+1)/n} H_s(x)\, dx - \frac{1}{n} H_s\left(\frac{i+1}{n}\right)\right\} \left\{F(b_1) - F(b_{i+1}) + \frac{i+1}{n} - \frac{1}{n}\right\}$$

$$(17) \qquad + \sqrt{n}\left\{\frac{1}{n} G_{n-1} - \int_0^{(n-1)/n} H_s(x)\, dx\right\}\left\{F(b_1) - F(b_n) + \frac{n-1}{n}\right\}$$

$$|A_2| \leqq n\, \frac{\sqrt{n}}{n}\, \underset{i}{\text{l.u.b.}} \max_{\frac{i}{n} \leqq x \leqq \frac{i+1}{n}} \left| H_s(x) - H_s\left(\frac{i+1}{n}\right)\right| \underset{i}{\text{l.u.b.}} \left|F(b_i) - \frac{i}{n}\right| + O(\sqrt{n}\left\{ |1 - F(b_n)| + \frac{1}{n} + F(b_1) \right\}.$$

The last term clearly goes to zero in probability since $n^{1-\eta}F(b_1)$ converges to zero in probability for $\eta > 0$. (The distribution function of $F(b_1)$ is $1 - (1 - x)^n$ for $0 \leqq x \leqq 1$.) By the Kolmogoroff theorem, given $\epsilon > 0$, we may choose A and N_1 such that for $n > N_1$

$$P(\text{l.u.b.}_i\, |F(b_i) - i/n|\, \sqrt{n} > A) < \epsilon.$$

Since $H_s(x)$ is uniformly continuous, we may choose N_2 so that for $n > N_2$

$$\underset{i}{\text{l.u.b.}} \max_{\frac{i}{n} \leqq x \leqq \frac{i+1}{n}} \left| H_s(x) - H_s\left(\frac{i+1}{n}\right)\right| \leqq \frac{\epsilon}{A}.$$

For $n > \max(N_1, N_2)$ the probability is less than ϵ that the first term on the right of (17) exceeds ϵ.

$$|A_4| \leqq \sqrt{n}\, \underset{i}{\text{l.u.b.}} \left|F(b_i) - \frac{i}{n}\right|^2 \sum_1^n \left\{\int_0^{i/n} H_s(x)\, dx\right\} \{F(b_{i+1}) - F(b_i)\}\, \frac{1}{F(b_{i+1})}\, \frac{n}{i+1}$$

$$\leqq \sqrt{n}\, \underset{i}{\text{l.u.b.}} \left|F(b_i) - \frac{i}{n}\right|^2 \max_x H_s(x) \sum_1^n \frac{F(b_{i+1}) - F(b_i)}{F(b_{i+1})}$$

$$\leqq C\sqrt{n}\ \underset{i}{\text{l.u.b.}}\left|F(b_i) - \frac{i}{n}\right|^2 \int_{F(b_1)}^1 \frac{dx}{x}$$

$$= C\sqrt{n}\ \underset{i}{\text{l.u.b.}}\left|F(b_i) - \frac{i}{n}\right|^2 \quad |\ln F(b_1)|.$$

By the usual argument $|A_4| \to 0$ in probability. We now return to A_3.

$$A_3 = -\frac{\sqrt{n}}{n}\sum_1^n \frac{n}{i+1}\int_0^{i/n} H_s(x)\,dx\left\{F(b_{i+1}) - \frac{i+1}{n}\right\}$$

$$+ \sqrt{n}\sum_1^n \frac{n}{i+1}\int_0^{i/n} H_s(x)\,dx\left\{F(b_i) - F(b_{i+1}) + \frac{1}{n}\right\}\ \left\{F(b_{i+1}) - \frac{i+1}{n}\right\}$$

$$= -\frac{1}{\sqrt{n}}\sum_1^n H_{s+1}\left(\frac{i}{n}\right)\left\{F(b_i) - \frac{i}{n}\right\}$$

$$(18) \qquad + \frac{1}{\sqrt{n}}\sum_1^{n-1}\frac{n}{i+1}\int_{i/n}^{(i+1)/n} H_s(x)\,dx\left\{F(b_{i+1}) - \frac{i+1}{n}\right\}$$

$$+ \frac{1}{\sqrt{n}}\frac{n}{n+1}\int_0^1 H_s(x)\,dx\cdot\frac{1}{n} + \frac{1}{\sqrt{n}}\,n\int_0^{1/n} H_s(x)\,dx\left\{F(b_1) - \frac{1}{n}\right\}$$

$$+ \sqrt{n}\sum_1^n \frac{n}{i+1}\int_0^{i/n} H_s(x)\,dx\left\{F(b_i) - F(b_{i+1}) + \frac{1}{n}\right\}\ \left\{F(b_{i+1}) - \frac{i+1}{n}\right\}$$

$$= \sum_1^5 B_i\,.$$

To complete the proof of Lemma 3 it is sufficient to show that $|B_i| \to 0$ in probability for $i = 2, 3, 4, 5$. The simplest arguments suffice for $i = 2, 3, 4$. In order to treat B_5 note the identity

$$2\sum (a_i - a_{i+1})a_{i+1} = a_1^2 - a_{K+1}^2 - \sum (a_i - a_{i+1})^2,$$

where sums are from 1 to K. If we set $a_i = F(b_i) - i/n$, we obtain

$$|2\sum \{F(b_i) - F(b_{i+1}) + 1/n\} \quad \{F(b_i) - i/n\}|$$

$$\leqq |F(b_1) - 1/n|^2 + |F(b_{K+1}) - (K+1)/n|^2 + \sum \{F(b_i) - F(b_{i+1}) + 1/n\}^2$$

$$\leqq 2\ \text{l.u.b.}_i\,|F(b_i) - i/n|^2 + \sum\{F(b_i) - F(b_{i+1})\}^2 + 3/n.$$

From the joint distribution of the quantities $\{F(b_i) - F(b_{i+1})\}$ given in [3] it is simple to show that

$$P\{F(b_{i+1}) - F(b_i) \geqq h\} = [1-h]^n, \qquad i = 1, 2, \cdots, n.$$

Therefore

$$P\ \{\underset{1<i<n}{\text{l.u.b.}}\ |F(b_{i+1}) - F(b_i)| < h\} = 1 - P(\cup_1^n\ |F(b_{i+1}) - F(b_i)| \geqq h)$$

$$\geqq 1 - \sum_1^n [1-h]^n = 1 - n[1-h]^n.$$

From this it follows that if $-1 < \alpha < 0$

$$\lim_{n\to\infty} P\{ \underset{1\leq i\leq n}{\text{l.u.b.}} |F(b_{i+1}) - F(b_i)| < n^{\alpha}\} = 1,$$

so that, choosing $\alpha = -7/8$, there exists an N such that, for $n > N$, with probability greater than $1 - \epsilon$

$$|\sum \{F(b_i) - F(b_{i+1}) + 1/n\} \quad \{F(b_{i+1}) - i/n\}| \leq \text{l.u.b.}_i |F(b_i) - i/n|^2 + n^{-3/4} + 3/2n.$$

Now applying the Abel transformation to B_5,

$$|B_5| \leq \sqrt{n}\left[\underset{i}{\text{l.u.b.}} \left|F(b_i) - \frac{i}{n}\right|^2 + \frac{1}{n^{3/4}} + \frac{1}{n}\right]$$

$$\cdot \sum_1^{n-1} \left|\frac{n}{i+1}\int_0^{i/n} H_s(x)\,dx - \frac{n}{i+2}\int_0^{(i+1)/n} H_s(x)\,dx\right| \tag{19}$$

$$+ \sqrt{n}\left[\underset{i}{\text{l.u.b.}} \left|F(b_i) - \frac{i}{n}\right|^2 + \frac{1}{n^{3/4}} + \frac{3}{2n}\right]\frac{n}{n+1}\int_0^1 H_s(x)\,dx.$$

The second term on the right clearly converges to zero in probability. We observe that

$$x^{-1}\int_0^x H_s(t)\,dt = \int_0^x t^{-1}[H_s(t) - H_{s+1}(t)]\,dt$$

and that by Lemma 2 the right side converges absolutely. Therefore

$$\sum_1^{n-1} \left|\frac{n}{i}\int_0^{i/n} H_s(x)\,dx - \frac{n}{i+1}\int_0^{(i+1)/n} H_s(x)\,dx\right|$$

$$= \sum_1^{n-1} \left|\int_0^{i/n} \frac{1}{x}[H_s - H_{s+1}]\,dx - \int_0^{(i+1)/n} \frac{1}{x}[H_s - H_{s+1}]\,dx\right|$$

$$= \sum_1^{n-1} \left|\int_{i/n}^{(i+1)/n} \frac{1}{x}[H_s - H_{s+1}]\,dx\right| \leq \int_0^1 \frac{1}{x}\{|H_s| + |H_{s+1}|\}\,dx = C_1.$$

Now

$$\sum_1^{n-1} \left|\frac{n}{i+1}\int_0^{i/n} H_s\,dx - \frac{n}{i+2}\int_0^{(i+1)/n} H_s\,dx\right|$$

$$\leq \sum_1^{n-1} \left|\frac{n}{i+1}\int_0^{(i+1)/n} H_s\,dx - \frac{n}{i+2}\int_0^{(i+2)/n} H_s\,dx\right|$$

$$+ \sum_1^{n-1} \frac{n}{i+1}\int_{i/n}^{(i+1)/n} H_s\,dx + \sum_1^{n-1} \frac{n}{i+2}\int_{(i+1)/n}^{(i+2)/n} H_s\,dx$$

$$\leq C_1 + 2\max_{0\leq x\leq 1} H_s(x) \sum_1^{n-1} \frac{1}{i} \leq C_1 + C_2 \ln n.$$

With this estimate the first term in (19) converges to zero in probability and Lemma 3 is completely proved.

4. Proof of main result. In statistics the random variable $F(b_{i+1}) - F(b_i)$ are known as coverages and the random variables

$$\{[F(b_{i+1}) - F(b_i)]/F(b_{i+1})\}, \qquad i = 0, 1, 2, \cdots, n,$$

are independent (as usual $F(b_0) = 0$ and $F(b_{n+1}) = 1$) with density $i(1-u)^{i-1}\,du$. From this one calculates easily that the independent random variables

$$X_{i,n} = \left\{\frac{F(b_i) - F(b_{i+1})}{F(b_{i+1})}\frac{i+1}{n} + \frac{1}{n}\right\}$$

have mean 0, variance $n^{-2}[i/(2+i)]$, and $E(|X_{i,n}|^3) \leqq 16n^{-3}$. In what follows it will be necessary to apply the Central Limit Theorem to sums of the form $S_n = \sum_{i=1}^n a_{i,n}X_{i,n}$. Although the Liapunoff version of the Central Limit Theorem is usually stated for sums of the form $(\sum \sigma_i^2)^{-1/2}\sum Y_i$, in this slightly more general form the proof [1] goes through with no change at all. It will be easy to show that the Liapunoff condition,

$$\lim_{n\to\infty}\frac{\sum |a_{i,n}|^3 E(|X_{i,n}|^3)}{\sum |a_{i,n}|^2 E(X_{i,n}^2))^{3/2}} = 0, \tag{20}$$

holds. Let

$$Y_{k,n} = \frac{1}{\sqrt{n}}\sum_{i=1}^n H_k(i/n)\left[F(b_i) - \frac{i}{n}\right], \qquad W_{k,n} = \sqrt{n}\sum_{i=1}^n \int_0^{i/n} H_k(x)\,dx\,X_{i,n},$$

$$V_n = \sqrt{n}\sum_{i=1}^n\left\{\int_0^{i/n} H_0(x)\,dx - \int_0^{i/n}\frac{dx}{x^2}\int_0^x tH_0(t)\,dt\right\}X_{i,n}.$$

Theorem 1 shows that we can replace $n^{-1/2}\sum_1^n F'(b_i)[F(b_i) - i/n]$ by V_n which is the sum of independent random variables.

THEOREM 1. *For every* $\epsilon > 0$

$$\lim_{n\to\infty} P\left[\left|n^{-1/2}\sum_1^n F'(b_i)[F(b_i) - i/n] - V_n\right| > \epsilon\right] = 0.$$

PROOF. By Lemma 1 it is sufficient to show that

$$\lim_{n\to\infty} P[|Y_{0,n} - V_n| > \epsilon] = 0. \tag{21}$$

This probability is

$$\begin{aligned} &P[|Y_{0,n} - V_n)| > \epsilon] \\ &\quad = P\Bigg[\Bigg|\sum_{k=0}^M(-1)^k\{Y_{k,n} - W_{k,n} + Y_{k+1,n}\} + \sum_{k=0}^M(-1)^k W_{k,n} - V_n \\ &\qquad\qquad + (-1)^{M+1}Y_{M+1,n}\Bigg| > \epsilon\Bigg] \\ &\quad \leqq \sum_{k=0}^M P[|Y_{k,n} - W_{k,n} + Y_{k+1,n}| > \epsilon/3M] \\ &\qquad + P\left[\left|\sum_{k=0}^M(-1)^k W_{k,n} - V_n\right| < \epsilon/3\right] + P[|Y_{M+1,n}| > \epsilon/3]. \end{aligned} \tag{22}$$

The sum in the second term can be evaluated as

$$\sum_{k=0}^{M} (-1)^k W_{k,n} - V_n$$

$$= \sqrt{n} \sum_{i=1}^{n} \left\{ \int_0^{i/n} \sum_{k=0}^{M} (-1)^k H_k(x)\, dx - \int_0^{i/n} H_0(x)\, dx + \int_0^{i/n} \frac{dx}{x^2} \int_0^x tH_0(t)\, dt \right\} X_{i,n}$$

$$= \sqrt{n} \left\{ \sum_{i=1}^{n} \int_0^{i/n} \left[\sum_{k=1}^{M} (-1)^k H_k(x)\, dx + \frac{1}{x^2} \int_0^x tH_0(t)\, dt \right] dx \right\} X_{i,n}.$$

By Lemma 2 the integrals appearing in the sum may be made uniformly less than $\sqrt{\epsilon^3/27}$ for all $M \geqq M_1$. The variance of $\sum (-1)^k W_{k,n} - V_n$ is then $\leqq \epsilon^3/27$ for all n.

By the Tchebycheff inequality

$$P\left[\left| \sum_{k=0}^{M} (-1)^k W_{k,n} - V_n \right| > \frac{\epsilon}{3} \right] \leqq \frac{\epsilon}{3}, \quad \text{all } n, \text{all } M \geqq M_1; \tag{23}$$

$$|Y_{M+1,n}| \leqq \sqrt{n}\ \text{l.u.b.}_i\ |F(b_i) - i/n|\ \max_x H_{M+1}(x). \tag{24}$$

Since $\lim_{k\to\infty} H_k(x) = 0$ uniformly, we can, by the Kolmogoroff theorem, find an M_2 and N_1 such that for $M \geqq M_2$ and $n > N_1$

$$P(|Y_{M+1,n}| > \epsilon/3) \leqq \epsilon/3. \tag{25}$$

Now fix an M, say $M = \max(M_1, M_2)$. Lemma 3 states that

$$\lim_{n\to\infty} P[|Y_{k,n} - W_{k,n} + Y_{k+1,n}| > \epsilon] = 0$$

for all k and all $\epsilon > 0$. Therefore for $N > N_2$

$$\sum P[|Y_{k,n} - W_{k,n} + Y_{k+1,n}| > \epsilon/3M] \leqq \epsilon/3. \tag{26}$$

Combining (22), (23), (25), and (26) we obtain $P[|Y_{0,n} - V_n| > \epsilon] \leqq \epsilon$ for all $n > \max(N_1, N_2)$, which proves Theorem 1.

It is now easy to prove Theorem 2, our main result, as stated at the outset. The variance of V_n, which is the sum of independent random variables, is asymptotically for large n

$$\frac{1}{n} \sum_{i=1}^{n} \left[\int_0^{i/n} H_0(t)\, dt - \int_0^{i/n} \frac{dx}{x^2} \int_0^x t\, H_0(t)\, dt \right]^2$$

$$\to \int_0^1 \left| \int_0^y H_0(t)\, dt - \int_0^y \frac{dx}{x^2} \int_0^x tH_0(t)\, dt \right|^2 dy = \nu^2.$$

It is easily verified that the Liapunoff condition (20) is satisfied, so that by the Central Limit Theorem

$$\lim_{n\to\infty} P(V_n < x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x/\nu} e^{-t^2/2}\, dt.$$

By Theorem 1,

$$\lim_{n\to\infty} P\left(\frac{1}{\sqrt{n}}\sum_1^n F'(b_i)\left[F(b_i) - \frac{i}{n}\right] < x\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x/\nu} e^{-t^2/2}\, dt. \tag{27}$$

Returning to (10),

$$\lim_{n\to\infty} \frac{1}{2n^{3/2}}\sum_1^n F'(b_i) = \lim_{n\to\infty}\frac{1}{2n^2}\sum_1^n F''(b_i) = 0; \frac{1}{n}\sum_1^n \left[F(b_i) - \frac{i}{n}\right] F''(b_i) \to 0$$

in probability.

By the weak law of large numbers, $n^{-1}\sum_1^n F'(b_i)$ converges in probability to $\int_{-\infty}^{\infty} F'^2(x)\, dF(x) = \mu$. Combining these with (9), (10), and (27) gives our final result,

$$\lim_{n\to\infty} P(\sqrt{n}\, a_n < x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{(\mu/\nu)x} e^{-t^2/2}\, dt.$$

The constants ν and μ cannot be zero. This is obvious for μ. If ν were zero, it would be necessary to have $\int_0^y H_0(t)\, dt = \int_0^y x^{-2}\, dx \int_0^x tH_0(t)\, dt$. Differentiating this twice leads to the equation $H_0(t) = c/t$, which is impossible.

REFERENCES

[1] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946, pp. 215–216.

[2] A. Kolmogoroff, "Confidence limits for an unknown distribution function," *Ann. Math. Stat.*, Vol. 12, (1941), pp. 461–463.

[3] S. Malmquist, "On a property of order statistics from a rectangular distribution," *Skand. Aktuarietids*, Vol. 33 (1950), pp. 214–222.

THE EXTREMA OF THE EXPECTED VALUE OF A FUNCTION OF INDEPENDENT RANDOM VARIABLES[1]

By Wassily Hoeffding

University of North Carolina

Summary. The problem is considered of determining the least upper (or greatest lower) bound for the expected value $EK(X_1, \cdots, X_n)$ of a given function K of n random variables $X_1, \cdots, X_n$ under the assumption that $X_1, \cdots, X_n$ are independent and each X_j has given range and satisfies k conditions of the form $Eg_i^{(j)}(X_j) = c_{ij}$ for $i = 1, \cdots, k$. It is shown that under general conditions we need consider only discrete random variables X_j which take on at most $k + 1$ values.

1. Introduction. Let $\mathcal{C}$ be the class of n-dimensional dfs (distribution functions) $F(\mathbf{x}) = F(x_1, \cdots, x_n)$ which satisfy the conditions

$$F(x_1, \cdots, x_n) = F_1(x_1)F_2(x_2) \cdots F_n(x_n), \tag{1.1}$$

$$\int g_i^{(j)}(x)\, dF_j(x) = c_{ij}, \qquad i = 1, \cdots, k; j = 1, \cdots, n, \tag{1.2}$$

$$F_j(x) = \begin{cases} 0 & \text{if } x < A_j, \\ 1 & \text{if } x > B_j, \end{cases} \qquad j = 1, \cdots, n, \tag{1.3}$$

where the functions $g_i^{(j)}(x)$ and the constants c_{ij}, A_j, and B_j are given. We allow that $A_j = -\infty$ and/or $B_j = \infty$. Here and in what follows, when the domain of integration is not indicated, the integral extends over the entire range of the variables involved. It will be understood that all dfs are continuous on the right.

Let $K(\mathbf{x})$ be a function such that

$$\phi(F) = \int K(\mathbf{x})\, dF(\mathbf{x})$$

exists and is finite for all F in $\mathcal{C}$. The problem is to determine the least upper and the greatest lower bound of $\phi(F)$ when F is in $\mathcal{C}$. It will be sufficient to consider only the least upper bound.

Special cases of statistical interest include $K(\mathbf{x}) = 0$ or 1 according as a function $f(\mathbf{x})$ does or does not exceed a given constant; $K(\mathbf{x}) = \max(x_1, \cdots, x_n)$, etc.; $g_i^{(j)}(x) = x^i$ (given moments up to order k); $g_i^{(j)}(x) = 1$ or 0 according as

Received November 3, 1953, revised November 30, 1954.

[1] This research was supported by the United States Air Force through the Office of Scientific Research of the Air Research and Development Command.

$x < b_i$ or $x \geqq b_i$ (given quantiles); etc. For

$$n = 1, \quad g_i^{(1)}(x) = x^i, \quad K(x) = \begin{cases} 1 \text{ if } x \leqq t, \\ K(x) = 0 \text{ otherwise} \end{cases}$$

the problem was stated and its solution found by Tchebycheff. This result was extended to more general function $K(x)$ by Markov, Possé, and others. References and proofs are given by Shohat and Tamarkin [9]. An extension to the case $g_i^{(1)}(x) = x^{m_i}$ and $A_1 = 0$ was considered by Wald [10]. Recent contributions are due to Karlin and Shapley [7] and Royden [8].

For n arbitrary, $k = 1$, $g_i^{(j)} = x$, $A_j = 0$, $B_j = \infty$, and $K(x) = 1$ or 0 according as $\sum_1^n x_i \geqq t$ or $<t$, Birnbaum, Raymond, and Zuckerman [3] showed that when looking for the least upper bound of $\phi(F)$ we need consider only dfs F_j which are step-functions with at most two steps. They gave an explicit solution for $n = 2$. For the case $k = 3$, $g_i^{(j)}(x) = x^i$ for $i = 1, 2$, and $g_3^{(j)}(x) = |x - c_{1j}|^3$, with $\phi(F)$ the distribution function of the sum $\sum_1^n x_i$, the inequalities of Berry [2], Esseen [4], and Bergström [1] give bounds which are asymptotically best as $n \to \infty$ but can presumably be improved for finite n.

In the present paper it is shown that if, in the general problem as stated above, we restrict ourselves to the subclass $\mathfrak{C}^*$ of $\mathfrak{C}$ where the F_j are step-functions with a finite number of steps, then we need consider only step-functions with at most $k + 1$ steps (Theorem 2.1); the number $k + 1$ can in general not be reduced. The same result holds in the unrestricted problem if (A) each F in $\mathfrak{C}$ can be approximated (in a certain sense) by a step-function in $\mathfrak{C}^*$, and (B) $\phi(F)$ is, in a sense, a continuous function of F (Theorem 2.2). Sufficient conditions for the fulfillment of assumptions (A) and (B) are given in Sections 3 and 4.

2. The main theorems. Denote by $\mathfrak{C}^*$ the class of all $F(\mathbf{x}) = F_1(x_1) \cdots F_n(x_n)$ in $\mathfrak{C}$ such that $F_1, \cdots, F_n$ are step-functions with a finite number of steps. Let $\mathfrak{C}_m$ be the subclass of $\mathfrak{C}^*$ in which each F_j for $j = 1, \cdots, n$ is a step-function with at most m steps.

THEOREM 2.1.

$$\sup_{F \in \mathfrak{C}^*} \phi(F) = \sup_{F \in \mathfrak{C}_{k+1}} \phi(F).$$

PROOF. Let $F(\mathbf{x}) = F_1(x_1) \cdots F_n(x_n)$ be an arbitrary df in $\mathfrak{C}^*$ such that for some j, F_j has more than $k + 1$ steps. It is sufficient to show that there exists a df G in $\mathfrak{C}_{k+1}$ such that $\phi(G) \geqq \phi(F)$. This, in turn, will easily follow when we show that if $F_n(x)$ has $m > k + 1$ steps, there exists a df $H_n(x)$ such that

a) $H_n(x)$ has less than m steps;
b) $H(\mathbf{x}) = F_1(x_1) \cdots F_{n-1}(x_{n-1})H_n(x_n)$ is in $\mathfrak{C}$;
c) $\phi(H) \geqq \phi(F)$.

By assumption $F_n(x)$ is of the form

$$F_n(x) = \begin{cases} 0 & \text{if} \quad x < a_1\,; \\ p_1 + \cdots + p_r & \text{if } a_r \leqq x < a_{r+1}, \quad r = 1, \cdots, m-1; \\ 1 & \text{if } a_m \leqq x; \end{cases}$$

where

$$A_n \leqq a_1 < a_2 < \cdots < a_m \leqq B_n\,;$$

$$p_r > 0, \qquad r = 1, \cdots, m;$$

$$g_{i1}p_1 + g_{i2}p_2 + \cdots + g_{im}p_m = c_{in}\,, \quad i = 0, 1, \cdots, k;$$

$$c_{0n} = 1; \quad g_{0r} = 1, \qquad r = 1, \cdots, m;$$

$$g_{ir} = g_i^{(n)}(a_r), \quad i = 1, \cdots, k; \quad r = 1, \cdots, m.$$

Let

$$H_n(x) = \begin{cases} 0 & \text{if} \quad x < a_1\,; \\ (p_1 + td_1) + \cdots + (p_r + td_r) & \\ & \text{if } a_r \leqq x < a_{r+1}\,; \quad r = 1, \cdots, m-1; \\ 1 & \text{if } a_m \leqq x. \end{cases}$$

In order to satisfy condition b) it is sufficient to choose t, and $d_1, \cdots, d_r$ in such a way that

$$p_r + td_r \geqq 0, \qquad r = 1, \cdots, m; \tag{2.1}$$

$$g_{i1}d_1 + g_{i2}d_2 + \cdots + g_{im}d_m = 0, \quad i = 0, 1, \cdots, k. \tag{2.2}$$

We can write $\phi(H) - \phi(F) = t\sum_{r=1}^{m} K_r d_r$ where

$$K_r = \int K(x_1, \cdots, x_{n-1}, a_r) \quad d\left\{\prod_{j=1}^{n-1} F_j(x_j)\right\}.$$

Let $\lambda = 0$ or 1 according as the rank of the matrix

$$\begin{matrix} g_{01}, \cdots, g_{0m} \\ \cdots \cdots \cdots \\ g_{k1}, \cdots, g_{km} \\ K_1, \cdots, K_m \end{matrix}$$

is less than or equal to $k + 2$. Then the equations (2.2) and $\sum_1^m K_r d_r = \lambda$ have a solution $(d_1, \cdots, d_m) \neq (0, \cdots, 0)$. Since $\sum_1^m d_r = 0$, at least one component of the solution $d_1, \cdots, d_m$ must be negative. Having thus fixed $d_1, \cdots, d_m$, we choose t as the largest number which satisfies the inequalities (2.1). This number exists and is positive. Conditions a), b) and c) are now satisfied, and the proof is complete.

Let the distance $d(F, G)$ between two dfs $F(\mathbf{x}) = F_1(x_1) \cdots F_n(x_n)$ and $G(\mathbf{x}) = G_1(x_1) \cdots G_n(x_n)$ be defined by

$$d(F, G) = \max_{1 \le j \le n} \sup_{-\infty < x < \infty} |F_j(x) - G_j(x)|.$$

We shall make the following assumptions.

Assumption A. For every F in $\mathfrak{C}$ and every $\delta > 0$ there exists an F^* in $\mathfrak{C}^*$ such that $d(F, F^*) < \delta$.

Assumption B. For every F in $\mathfrak{C}$ and every $\epsilon > 0$ there exists $\delta > 0$ such that for any G in $\mathfrak{C}$ which satisfies $d(F, G) < \delta$ we have $|\phi(F) - \phi(G)| < \epsilon$.

Lemma 2.1. *If Assumptions* A *and* B *are satisfied,*

$$\sup_{F \in \mathfrak{C}} \phi(F) = \sup_{F \in \mathfrak{C}^*} \phi(F).$$

The proof is obvious. An analogous theorem clearly holds with an arbitrary distance function.

From Theorem 2.1 and Lemma 2.1 we obtain

Theorem 2.2. *If Assumptions* A *and* B *are satisfied,*

$$\sup_{F \in \mathfrak{C}} \phi(F) = \sup_{F \in \mathfrak{C}_{k+1}} \phi(F).$$

In Sections 3 and 4 it will be shown that Assumptions A and B are satisfied for certain classes $\mathfrak{C}$ and functions K which are of interest in statistics.

We conclude this section by showing that Theorem 2.1 cannot be improved without imposing additional restrictions. More precisely, for every k and every n there exist functions K, $g_i^{(j)}$ and constants A_j, B_j, c_{ij} such that the conditions of the theorem are satisfied and $\sup_{\mathfrak{C}} \phi(F) > \sup_{\mathfrak{C}_m} \phi(F)$ if $m < k + 1$. Furthermore, the functions $g_i^{(j)}$ and the constants A_j, B_j, c_{ij} can be chosen to be independent of j. Let

$$g_i^{(j)}(x) = x^i, \qquad c_{ij} = c_i, \qquad A_j = A, \qquad B_j = B, \qquad j = 1, \cdots, n,$$

where A and B are finite.

First assume $n = 1$. We can choose the constants c_i and $k + 1$ distinct real numbers $a_1, a_2, \cdots, a_{k+1}$ in $[A, B]$ in such a way that the equations

$$a_1^i p_1 + a_2^i p_2 + \cdots + a_{k+1}^i p_{k+1} = c_i, \qquad i = 0, 1, \cdots, k,$$

where $c_0 = 1$, have a unique solution $(p_1, \cdots, p_{k+1})$ with $p_r > 0$ for all r (see, for example, [9]). Then the df F_0, which assigns probability p_r to the point a_r for all r, is in $\mathfrak{C}_{k+1}$. Let $K_1(x)$ be a nonnegative continuous function, and let $Q(x)$ be a polynomial of degree k such that $K_1(x) \le Q(x)$ for $A \le x \le B$ and equality holds if and only if $x = a_r$ for $r = 1, 2, \cdots, k + 1$. Let $\phi_1(F) = \int K_1(x)\, dF(x)$. Then for every F in $\mathfrak{C}$,

$$\phi_1(F) \le \int Q(x)\, dF(x) = \phi_1(F_0),$$

and equality holds only for $F = F_0$.

Now suppose that $\sup_{\mathfrak{C}_m} \phi_1(F) = \phi_1(F_0)$ for some $m < k + 1$. Then there exists a sequence $\{F^{(s)}\}$ of cdfs in $\mathfrak{C}_m$ such that $\lim \phi_1(F^{(s)}) = \phi_1(F_0)$. Since A and B are finite, there exists, by the Bolzano-Weierstrass theorem, a subsequence $\{F^{(s_j)}\}$ of $\{F^{(s)}\}$ which converges to a function F^* in $\mathfrak{C}_m$ at all points of continuity of F^*. Since $K_1(x)$ is continuous, this implies $\lim \phi_1(F^{(s_j)}) = \phi_1(F^*) = \phi_1(F_0)$. This is a contradiction since $F^* \neq F_0$.

For n arbitrary let $K(x_1, \cdots, x_n) = K_1(x_1)K_1(x_2) \cdots K_1(x_n)$, so that $\phi(F) = \phi_1(F_1) \cdots \phi_1(F_n)$. If the conditions for the case $n = 1$ are satisfied, we arrive at the desired conclusion, making use of the condition $K_1(x) \geqq 0$.

The assumptions that K_1 is continuous and A, B are finite are inessential, at least for $k = 2$. This is seen from the fact that the bound of the Bienaymé-Tchebycheff inequality can not in general be arbitrarily closely approached when the distribution is a two-step-function.

3. Approximation of a df by a step-function. It will now be shown that Assumption A is satisfied if $\mathfrak{C}$ is the class of distributions of the product type (1.1) with prescribed moments and ranges.

THEOREM 3.1. *Assumption* A *is satisfied if* $\mathfrak{C}$ *is the class defined by* (1.1) *to* (1.3) *with* $g_i^{(j)}(x) = x^{m_{ij}}$, *where the* m_{ij} *are arbitrary positive integers.*

The theorem is an immediate consequence of

LEMMA 3.1. *Let* $F(x)$ *be a df on the real line such that*

$$\int x^i \, dF(x) = c_i, \qquad i = 1, \cdots, s; \tag{3.1}$$

$$F(x) = \begin{cases} 0 & \text{if } x < A, \\ 1 & \text{if } x > B, \end{cases} \tag{3.2}$$

where we may have $A = -\infty$ *or* $B = \infty$. *Then for every* $\delta > 0$ *there exists a cdf* $F^*(x)$ *which is a step-function with a finite number of steps, satisfies conditions* (3.1) *and* (3.2), *and for which*

$$\sup_x |F^*(x) - F(x)| < \delta.$$

To prove Lemma 3.1 we shall need

LEMMA 3.2. *If* $F(x)$ *is any df which satisfies conditions* (3.1) *and* (3.2), *there exists a df which is a step-function with a finite number of steps and satisfies the same conditions.*

The statement of Lemma 3.2 is well known. For example, it follows from Shohat and Tamarkin ([9], Theorems 1.2 and 1.3 and Lemma 3.1).

PROOF OF LEMMA 3.1. Given $\delta > 0$, we can choose a finite set of points

$$A = a_0 < a_1 < a_2 < \cdots < a_m < a_{m+1} = B$$

such that $p_r = F(a_{r+1} - 0) - F(a_r) < \delta$ for $r = 0, 1, \cdots, m$. If $p_r \neq 0$, let

$$F_r(x) = \begin{cases} 0 & \text{if} \quad x < a_r \\ p_r^{-1}[F(x) - F(a_r)] & \text{if} \quad a_r \leqq x < a_{r+1} \\ 1 & \text{if } a_{r+1} \leqq x. \end{cases}$$

By Lemma 3.2 there exists a df $F_r^*(x)$ which is a step-function with a finite number of steps and such that

$$\int x^i \, dF_r^*(x) = \int x^i \, dF_r(x), \qquad i = 1, \cdots, s,$$

$$F_r^*(x) = \begin{cases} 0 & \text{if } x < a_r, \\ 1 & \text{if } x > a_{r+1}. \end{cases}$$

Let

$$F^*(x) = \begin{cases} 0 & \text{if} \quad x < A; \\ F(a_r) + p_r F_r^*(x) & \text{if } a_r \leqq x < a_{r+1}, \quad r = 0, 1, \cdots, m; \\ 1 & \text{if} \quad x \geqq B; \end{cases}$$

where $p_r F_r^*(x) = 0$ if $p_r = 0$. It can be verified that $F^*(x)$ has the properties stated in Lemma 3.1.

4. Continuity of $\phi(F)$. In this section we consider sufficient conditions for the continuity of $\phi(F)$ in the sense of Assumption B. The next theorem shows that Assumption B is satisfied if $\phi(F)$ is the probability that the random vector $\mathbf{X}$ with df F is contained in a set S of a fairly general type.

THEOREM 4.1. *Assumption* B *is satisfied if* $K(x) = 1$ *or* 0 *according as x does or does not belong to a Borel set S such that every set*

$$S_j(x_1, \cdots, x_{j-1}, x_{j+1}, \cdots, x_n) = \{x_j : \mathbf{x} \in S, x_h \text{ fixed for } h \neq j\}, \quad j = 1, \cdots, n,$$

is the union of a finite and bounded number of intervals.

(Here $\{x\colon C\}$ denotes the set of all points x which satisfy condition C.)

PROOF. Let δ be an arbitrary positive number. Let $F(\mathbf{x}) = F_1(x_1) \cdots F_n(x_n)$ and $G(\mathbf{x}) = G_1(x_1) \cdots G_n(x_n)$ be any two dfs in $\mathfrak{C}$ such that

$$\sup_x |F_h(x) - G_h(x)| < \delta, \qquad h = 1, \cdots, n.$$

Let

$$F^{(j)}(\mathbf{x}) = \left\{ \prod_{h=1}^{j} F_h(x_h) \right\} \left\{ \prod_{h=j+1}^{n} G_h(x_h) \right\}, \qquad j = 0, 1, \cdots, n,$$

so that $F^{(0)}(\mathbf{x}) = G(\mathbf{x})$ and $F^{(n)}(\mathbf{x}) = F(\mathbf{x})$. Then

$$\phi(F) - \phi(G) = \sum_{j=1}^{n} [\phi(F^{(j)}) - \phi(F^{(j-1)})].$$

We may assume that every set S_j is the union of at most N nonoverlapping intervals, where N is a fixed number. For a fixed integer j and fixed values x_h, with $h \neq j$, denote these intervals by $I_1, I_2, \cdots, I_M$, where $M \leqq N$. We have

$$\phi(F^{(j)}) - \phi(F^{(j-1)})$$

$$= \int L(x_1, \cdots, x_{j-1}, x_{j+1}, \cdots, x_n) \, d\{F_1(x_1) \cdots F_{j-1}(x_{j-1}) G_{j+1}(x_{j+1}) \cdots G_n(x_n)\},$$

where $L(x_1, \cdots, x_{j-1}, x_{j+1}, \cdots, x_n) = \sum_{m=1}^{M} \{\int_{I_m} dF_j(x_j) - \int_{I_m} dG_j(x_j)\}$. Since

$$\left| \int_{I_m} dF_j(x_j) - \int_{I_m} dG_j(x_j) \right| < 2\delta,$$

we get $|\phi(F) - \phi(G)| < 2nM\delta \leqq 2nN\delta$, so that Assumption B is satisfied.

Assumption B of Theorem 2.2 is evidently satisfied under more general conditions than those of Theorem 4.1. Thus if $K(\mathbf{x}) = d_1K_1(\mathbf{x}) + \cdots + d_RK_R(\mathbf{x})$, where $d_1, \cdots, d_R$ are arbitrary constants and $\phi_r(F) = \int K_r(\mathbf{x})\, dF(\mathbf{x})$ is continuous in the sense of Assumption B for $r = 1, \cdots, R$, then $\phi(F) = \int K(\mathbf{x})\, dF(\mathbf{x})$ is also continuous. The same is true if $K(\mathbf{x})$ can be approximated by a function of the form $\sum d_rK_r(\mathbf{x})$, uniformly in the range of the distributions in $\mathfrak{C}$.

Using Theorems 3.1 and 4.1 we can state the following corollary of Theorem 2.2.

THEOREM 4.2. *Let $\mathfrak{C}$ be the class of dfs $F(\mathbf{x}) = F_1(x_1) \cdots F_n(x_n)$ such that*

$$F_j(A_j - 0) = 0, \qquad F_j(B_j + 0) = 1, \qquad \int x^{m_{ij}}\, dF_j(x) = c_{ij},$$
$$i = 1, \cdots, k;\ j = 1, \cdots, n,$$

where the numbers A_j, B_j, c_{ij} and the integers m_{ij} are given. Let S be a Borel set such that every set $\{x_j : \mathbf{x} \in S, x_h \text{ fixed for } h \neq j\}$ is the union of a finite and bounded number of intervals. Then

$$\sup_{F \in \mathfrak{C}} \int_S dF = \sup_{F \in \mathfrak{C}_{k+1}} \int_S dF.$$

5. Concluding remark. The problem considered in this paper can be modified by admitting only those dfs in $\mathfrak{C}$ for which the marginal distributions $F_1, \cdots, F_n$ are identical. Some results for this case were obtained in [6]. With this restriction Theorem 2.1 is no longer true, and the assumptions are no longer sufficient in order to reduce the class of competing dfs to step-functions with a bounded number of steps.

For example, consider the problem of the least upper bound for the expected value of the largest of n independent, identically distributed random variables with given mean and variance. Hartley and David [5] showed that under the additional assumption that the df is continuous the least upper bound is attained with a continuous df when $n \geqq 2$. At least for $n = 2$ it can be shown [6] that the Hartley-David bound cannot be arbitrarily closely approached with a discrete df having a bounded number of steps.

On the other hand, if the assumption that the random variables are identically distributed is dropped, Theorem 2.2 implies that the least upper bound is attained or approached with step-functions having at most three steps.

REFERENCES

[1] H. Bergström, "On the central limit theorem in the case of not equally distributed random variables", *Skand. Aktuarietids.*, Vol. 32 (1949), pp. 37–62.

[2] A. C. Berry, "The accuracy of the Gaussian approximation to the sum of independent variates", *Trans. Amer. Math. Soc.*, Vol. 49 (1941), pp. 122–136.

[3] Z. W. Birnbaum, J. Raymond and H. S. Zuckerman, "A generalization of Tshebyshev's inequality to two dimensions", *Ann. Math. Stat.*, Vol. 18 (1947), pp. 70–79.

[4] C. G. Essen, "Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law", *Acta Math.*, Vol. 77, (1944), pp. 1–125.

[5] H. O. Hartley and H. A. David, "Universal bounds for mean range and extreme observation", *Ann. Math. Stat.*, Vol. 25 (1954), pp. 85–99.

[6] W. Hoeffding and S. S. Shrikhande, "Bounds for the distribution function of a sum of independent, identically distributed random variables", *Ann. Math. Stat.*, to be published.

[7] S. Karlin and L. S. Shapley, "Geometry of moment spaces", *Mem. Amer. Math. Soc.*, No. 12 (1953).

[8] H. L. Royden, "Bounds on a distribution function when its first n moments are given", *Ann. Math. Stat.*, Vol. 24 (1953), pp. 361–376.

[9] J. A. Shohat and J. D. Tamarkin, *The Problem of Moments*, Mathematical Surveys, No. II, American Mathematical Society, New York, 1943.

[10] A. Wald, "Limits of a distribution function determined by absolute moments and inequalities satisfied by absolute moments", *Trans. Amer. Math. Soc.*, Vol. 46 (1939), pp. 280–306.

ESTIMATES OF BOUNDED RELATIVE ERROR IN PARTICLE COUNTING[1]

By M. A. Girshick, H. Rubin, and R. Sitgreaves

Stanford University

Summary. A statistical problem arising in many fields of activity requires the estimation of the average number of events occurring per unit of a continuous variable, such as area or time. The underlying distribution of events is assumed to be Poisson; the constant to be estimated is the unknown parameter λ of the distribution.

A sampling procedure is proposed in which the continuous variable is observed until a fixed number M of events occurs. Such a procedure enables us to form an estimate l, which with confidence coefficient α does not differ from λ by more than 100 γ per cent of λ. The values of γ and α depend on M but not on λ.

Modifications of this procedure which are sequential in nature and have possible operational advantages are also described.

These procedures are discussed in terms of a chemical problem of particle counting. It is clear, however, that they are generally applicable whenever the basic probability assumptions apply.

1. Introduction. The following problem arising in chemical research is typical of a statistical problem occurring in many fields of activity. A set of inert particles is randomly distributed over a microscope slide of area A. It is assumed that the probability of m particles falling in a subset of area a is

$$P_\lambda(m \mid a) = e^{-\lambda a}(\lambda a)^m \,/\, m!$$

and that distributions in disjoint areas are independent. On the basis of particle counts in subareas of A, we want to estimate the unknown parameter λ. With this estimate we can, by performing the indicated multiplication, also estimate, if desired, $N = A\lambda$, the expected number of particles on the slide. In addition, we would like to make a confidence statement about the reliability of our estimate.

In general, in discussing the reliability of an estimate, we would like to bound, at a given confidence level, either the absolute or the relative error of the estimate. That is, for a selected estimate t of an unknown parameter θ, we would like to say with confidence coefficient α, either

$$|t - \theta| \leqq \beta \quad \text{or} \quad |t/\theta - 1| \leqq \gamma, \quad \text{i.e.,} \ (1 - \gamma) \leqq t/\theta \leqq (1 + \gamma),$$

where α and β (or γ) are suitably chosen numbers which do not depend on θ.

In the present problem, it seems reasonable to be concerned with the relative

Received November 18, 1952, revised May 20, 1954.

[1] This research was the result of a discussion of one of the authors with Dr. W. A. Perkins of the Chemistry Department, Stanford University.

rather than the absolute error of the estimate, since this error is independent of the units in which the area is measured and is the same whether we are estimating λ, or $A\lambda$, the expected number of particles on the slide. What we would like, therefore, is an estimate l of λ such that we can say with confidence coefficient α that l does not differ from λ by more than 100γ per cent of λ, where neither α nor γ depend on the true value of λ. Such an estimate we call an estimate of *bounded relative error*. Moreover, among all estimates of this type we would like to choose one possessing some optimal properties.

2. Fixed area sampling procedure. In estimating λ in problems of the present kind, it often has been the practice to select n non-overlapping subareas of A, each of size a_0, with n and a_0 determined in advance, and to count the particles in the selected subareas. The usual estimate of λ is then

$$l = \frac{1}{na_0} \sum_{i=1}^{n} x_i$$

where x_i is the number of particles observed in the ith subarea. It is easy to see, however, that for a given value of γ,

$$P\{\lambda(1-\gamma) \leqq l \leqq \lambda(1+\gamma)\}$$

is a function of the unknown parameter λ, so that no confidence statement concerning a bound on the relative error of l can be made.

On the other hand, from the observed value of l, we can determine values for two functions L_μ and L_ν defined by

$$\sum_{j=na_0\bar{l}}^{\infty} \frac{\exp\{-L_\mu(l)na_0\}(L_\mu(l)na_0)^j}{j!} = \frac{1-\alpha}{2},$$

$$\sum_{j=0}^{na_0\bar{l}} \frac{\exp\{-L_\nu(l)na_0\}(L_\nu(l)na_0)^j}{j!} = \frac{1-\alpha}{2}.$$

Then, we can say with confidence coefficient approximately α that

$$L_\mu(l) \leqq \lambda \leqq L_\nu(l),$$

$$\lambda[1-\gamma(l)] \leqq \hat{l}(l) \leqq \lambda[1+\gamma(l)],$$

where

$$l(l) = \frac{2L_\nu)l)L_\mu(l)}{L_\nu(l) + L_\mu(l)}, \qquad \gamma(l) = \frac{L_\nu(l) - L_\mu(l)}{L_\nu(l) + L_\mu(l)}.$$

It follows that if we estimate λ by $\hat{l}(l)$, we can say with confidence coefficient approximately α that the relative error of $\hat{l}(l)$ is bounded by $\gamma(l)$, but this quantity is a chance variable whose expected value depends on the unknown value of λ.

Since a fixed area sampling procedure does not yield estimates of bounded relative error, we consider the possibility of using another type of sampling

procedure. (Some properties of fixed count and fixed time experiments are discussed by Albert and Nelson [1] for the case of counter data for radioactivity.) By a suitable adjustment of the microscope, the aperture can be increased gradually so that the area under observation increases continuously from a point to any desired magnitude, within the limits of the area of the slide.

If the area is expanded only until a fixed number of particles is counted, the magnitude of the observed area becomes a chance variable with an attached probability distribution. In this distribution, the parameter λ appears as a scale parameter. It has been shown [2] that if the single unknown parameter of a distribution is a scale parameter, estimates of bounded relative error exist, and among them one possessing specified optimal properties can be found. By adopting a fixed particle count procedure and applying the general theory, our problem is formally solved.

In the following sections, a fixed particle count procedure is explicitly defined and an estimate of bounded relative error is proposed. For the given value of γ and number of particles counted, it maximizes the value of the confidence coefficient α. Some modifications of this procedure are also discussed.

3. Fixed particle count procedure. Suppose in counting particles under a microscope, the area under observation is expanded until a fixed number of particles, M, is counted. The magnitude of the area so obtained, say a_M, is a chance variable with an attached probability distribution. Since we have assumed that the probability of m particles falling in a subset of fixed area a is

$$P_\lambda(m \mid a) = (\lambda a)^m e^{-\lambda a} / m!,$$

the probability that an area of size x/λ has fewer than M particles is

$$P\{\lambda a_M > x\} = e^{-x} \sum_{j=0}^{M-1} \frac{x^j}{j!} = \int_x^\infty \frac{e^{-t} t^{M-1}}{(M-1)!}\, dt.$$

It follows that $\lambda a_M = X$ has the density function

$$f_M(x) = e^{-x} x^{M-1} / (M-1)!.$$

That is, λa_M has a gamma distribution with parameter M, or $2\lambda a_M$ has a chi-square distribution with $2M$ degrees of freedom.

Suppose now that an observation a_M is made, and we form the estimate l of λ defined by $l = b/a_M$, where b is a given positive number. If γ is the desired bound on the relative error, the probability, before the observation is made, that

$$(1 - \gamma) \leqq l/\lambda \leqq (1 + \gamma)$$

is given by

$$P\left\{\frac{b}{1+\gamma} \leqq \lambda a_M \leqq \frac{b}{1-\gamma}\right\} = \int_{b/1+\gamma}^{b/1-\gamma} \frac{x^{M-1} e^{-x}}{(M-1)!}\, dx = \psi_\gamma(b, M), \text{ say.}$$

The quantity $\psi_\gamma(b, M)$ does not depend upon λ, so that for all λ, l is an estimate with relative error bounded by γ at confidence level $\psi_\gamma(b, M)$. Clearly, we would like to find that value of b, say b^*, for which $\psi_\gamma(b, M)$ is a maximum.

As b varies from 0 to ∞, the value of $\psi_\gamma(b, M)$ increases continuously from zero to a maximum value $\psi_\gamma(b^*, M)$, and then decreases again to zero. To determine b^*, we consider

$$\frac{\partial}{\partial b}\psi_\gamma(b, M) = \frac{\partial}{\partial b}\left[\int_0^{b/1-\gamma} \frac{x^{M-1}e^{-x}}{(M-1)!}\,dx - \int_0^{b/1+\gamma} \frac{x^{M-1}e^{-x}}{(M-1)!}\,dx\right]$$
$$= \frac{b^{M-1}}{(M-1)!}\left[\frac{e^{-b/(1-\gamma)}}{(1-\gamma)^M} - \frac{e^{-b/(1+\gamma)}}{(1+\gamma)^M}\right].$$

The maximizing value is the single finite positive value of b for which

$$\partial\psi_\gamma(b, M)\,/\partial b = 0;$$

that is,

$$b^* = \frac{M(1-\gamma^2)}{2\gamma}\log\frac{1+\gamma}{1-\gamma}. \tag{3.1}$$

For b^*, we have

$$\psi_\gamma(b^*, M) = \int_{\mathfrak{D}} \frac{x^{M-1}e^{-x}}{(M-1)!}\,dx = \psi_\gamma^*(M), \text{ say,}$$

$$\mathfrak{D} = \left[\frac{M(1-\gamma)}{2\gamma}\log\frac{1+\gamma}{1-\gamma}, \frac{M(1+\gamma)}{2\gamma}\log\frac{1+\gamma}{1-\gamma}\right].$$

For a fixed value of γ, the function ψ_γ^* is a single-valued monotone increasing function. The monotonicity of ψ_γ^* follows from the fact that a_M is a sufficient statistic for $a_1, \cdots, a_M$. It follows, therefore, that if we define M_0 to be the least integer such that $\psi_\gamma^*(M) \geqq \alpha$, then $M_0 = \eta_\gamma(\alpha)$ is a single-valued monotone increasing function of α.

With these considerations in mind, we propose the following fixed particle count procedure. The desired values of γ and α are specified in advance, and from these we determine $M = \eta_\gamma(\alpha)$. The area under observation is expanded until M particles are counted, resulting in an observation a_M. We estimate λ, the average number of particles per unit area, and N, the expected number of particles on the slide with total area A, by

$$l^* = b^*\,/\,a_M, \qquad N^* = Al^*,$$

where b^* is defined in (3.1). In either case, the relative error of the estimate is bounded by γ at confidence level at least α, regardless of the true value of λ.

If our estimation problem is formulated as a decision function problem in which our action space is the positive half of the real line, that is,

$$\{l:\ 0 \leqq l < \infty\},$$

and our loss function is

$$L(\lambda, l) = \begin{cases} 0 & \text{if } |(l/\lambda) - 1| \leq \gamma, \\ 1 & \text{if } |(l/\lambda) - 1| > \gamma, \end{cases}$$

the proposed estimate l^* is the best invariant estimate. Since the loss function is bounded, it is also minimax [3]. The risk when we estimate λ by l^* is $1 - \psi_\gamma^*(M)$.

4. Values of γ, M, and $\psi_\gamma^*(M)$. Table 1 gives values of $\psi_\gamma^*(M)$ for four values of γ and for $M \leqq 40$. For $M > 40$, the distribution of $(\sqrt{4\lambda a_M} - \sqrt{4M - 1})$ is approximately normal with zero mean and unit variance. In this case, therefore,

$$\begin{aligned} \psi_\gamma^*(M) \sim \Phi\left(\sqrt{\frac{4M(1+\gamma)}{2\gamma}\log\frac{1+\gamma}{1-\gamma}} - \sqrt{4M-1}\right) \\ - \Phi\left(\sqrt{\frac{4M(1-\gamma)}{2\gamma}\log\frac{1+\gamma}{1-\gamma}} - \sqrt{4M-1}\right), \end{aligned} \tag{4.1}$$

where $\Phi(u) = \int_{-\infty}^{u} (2\pi)^{-\frac{1}{2}} e^{-t^2/2}\, dt$. Since $0 < \gamma < 1$, we have

$$\frac{1}{2\gamma}\log\frac{1+\gamma}{1-\gamma} = 1 + \frac{\gamma^2}{3} + \frac{\gamma^4}{5} + \frac{\gamma^6}{7} + \cdots$$

$$\psi_\gamma^*(M) \sim \Phi\left[\sqrt{4M}\left(\frac{\gamma^2}{24} + \frac{\gamma}{2} + \frac{7}{48}\gamma^3\right)\right] - \Phi\left[\sqrt{4M}\left(\frac{\gamma^2}{24} - \frac{\gamma}{2} - \frac{7}{48}\gamma^3\right)\right].$$

For values of γ and α which are generally of practical interest, a good approximation for the required value of M is given by the relation

$$\sqrt{4M}\left(\frac{\gamma}{2} + \frac{7}{48}\gamma^3\right) = z_\alpha \tag{4.2}$$

where $z_\alpha = \Phi^{-1}[\frac{1}{2}(1 + \alpha)]$, that is, $\Phi(z_\alpha) - \Phi(-z_\alpha) = \alpha$.

For $\gamma = .10$ and $\alpha = .90$, $.95$, and $.99$, respectively, this approximation yields the same values of M as those determined from (4.1), while for $\gamma = .05$, the values differ only slightly:

	α: .90	.95	.99	
$\gamma = .10$	269	382	660	(4.1) or (4.2)
$\gamma = .05$	1081	1535	2651	(4.1)
	1081	1534	2650	(4.2)

5. Modified fixed particle count procedure. Instead of expanding the area continuously until M particles are counted, it may be more convenient to adopt a sampling procedure consisting of k subsamples. In the jth subsample, the area is expanded continuously until a fixed number of particles m_j is counted, with $\sum_1^k m_j = M$, and with the provision that areas observed in the different sub-

TABLE 1

Values of $\psi_\gamma^*(M) = \frac{1}{(M-1)!}\int_{\mathfrak{D}} x^{M-1}e^{-x}\,dx$,

$$\mathfrak{D} = \left[\frac{M(1-\gamma)}{2\gamma}\log\frac{(1+\gamma)}{(1-\gamma)}, \frac{M(1+\gamma)}{2\gamma}\log\frac{(1+\gamma)}{(1-\gamma)}\right].$$

M	γ			
	.01	.05	.10	.20
2	.0108	.0541	.1083	.2165
4	.0156	.0781	.1558	.3084
6	.0193	.0962	.1915	.3753
8	.0223	.1114	.2211	.4292
10	.0250	.1247	.2469	.4746
12	.0274	.1367	.2700	.5140
14	.0297	.1476	.2909	.5487
16	.0317	.1578	.3102	.5796
18	.0337	.1674	.3281	.6075
20	.0355	.1763	.3449	.6329
22	.0373	.1849	.3607	.6560
24	.0389	.1930	.3757	.6771
26	.0405	.2008	.3898	.6966
28	.0421	.2082	.4033	.7146
30	.0436	.2154	.4162	.7313
32	.0450	.2223	.4286	.7468
34	.0464	.2290	.4404	.7612
36	.0477	.2355	.4518	.7746
38	.0490	.2418	.4628	.7871
40	.0503	.2479	.4733	.7988

samples are nonoverlapping. If $a_{m_j}^{(j)}$ is the area observed in the jth subsample, we now estimate λ by

$$l_k^* = b^* \Big/ \sum_{j=1}^{k} a_{m_j}^{(j)}.$$

Since each of the $a_{m_j}^{(j)}$ is independently distributed in a gamma distribution with parameter m_j, their sum, because of the reproductive property of the gamma (or chi-square) distribution, is again a gamma distribution with parameter $\sum_1^k m_j = M$. The theory now goes through as before.

6. Sequential Procedure. For practical purposes, it may be convenient to modify the sampling procedure in the following manner. The desired bound on the relative error, γ, and the confidence coefficient α are specified, and the corresponding value of M required for the fixed particle count procedure is de-

termined. The area under observation is now expanded until a fixed number of particles, r, is counted; $1 \leq r < M$.

Let a_r be the size of the area so observed. Subsequently, non-overlapping areas of fixed size ca_r, where c is a given positive number, are examined successively, and the number of particles in each area is counted. Let ν_i be the number of particles counted in the ith area of size ca_r. Sampling stops with the nth such area if

$$N' < M - r \leq N = M - r + u,$$

say, where

$$N' = \sum_{i=1}^{n-1} \nu_i, \qquad N = \sum_{i=1}^{n} \nu_i.$$

In this procedure, the quantities $a_r, \nu_1, \cdots, \nu_n$, and n are all chance variables. When sampling stops, we estimate λ by

$$l^* = \frac{(M+u)(1-\gamma^2)}{2\gamma a_r(1+nc)} \log \frac{1+\gamma}{1-\gamma}.$$

That is, we estimate λ in the same manner as before, but replace M and a_n by $(m + u)$ and $a_r(1 + nc)$, respectively. We state with a confidence coefficient at least as large as α that the relative error of l^* is bounded by γ.

This statement is justified as follows: the joint probability density of $\nu_1, \cdots, \nu_n$, n and $X = \lambda a_r$ is given by

$$\phi(\nu_1, \cdots, \nu_n, n, x) = \frac{e^{-x}x^{r-1}}{(r-1)!} \prod_{i=1}^{n} \left(\frac{e^{-cx}(cx)^{\nu_i}}{\nu_i!} \right),$$

$$\text{(6.1)} \qquad \sum_{i=1}^{n-1} \nu_i < M - r \leq \sum_{i=1}^{n} \nu_i.$$

(The density is with respect to discrete measure for $\nu_1, \cdots, \nu_n$, n, and to Lebesgue measure for x.) This can be written as

$$\phi(\nu_1, \cdots, \nu_n, n, x) = \frac{e^{-x(1+nc)}(1+nc)^{r+N}x^{r+N-1}}{(r+N-1)!} \cdot h(\nu_1, \cdots, \nu_n, n).$$

The probability before any observations are made that

$$1 - \gamma \leq \frac{l^*}{\lambda} \leq 1 + \gamma$$

is equal to

$$\sum_{n=1}^{\infty} \sum h(\nu_1, \cdots, \nu_n, n) P \left\{ \left[\frac{1-\gamma}{2\gamma} \log \frac{1+\gamma}{1-\gamma} \leq \frac{\lambda a_r(1+nc)}{r+N} \right.\right.$$
$$\left.\left. \leq \frac{1+\gamma}{2\gamma} \log \frac{1+\gamma}{1-\gamma} \right] \middle| \nu_1, \cdots, \nu_n, n \right\}$$

$$= \sum_{n=1}^{\infty} \sum h(\nu_1, \cdots, \nu_n, n) \cdot \psi_\gamma^*(r+N)$$

$$\geq \psi_\gamma^*(M) \sum_{n=1}^{\infty} \sum h(\nu_1, \cdots, \nu_n, n)$$

$$= \alpha.$$

All the sums without indices are over all sequences $\nu_1, \cdots, \nu_n$ satisfying (6.1).

7. Probability distribution of n. For any sequential procedure, it is of interest to determine, if possible, the probability distribution of n together with its expected value. For the proposed procedure, this distribution is given by

$$g(n) = \sum h(\nu_1, \cdots, \nu_n, n)$$

$$= \sum \frac{(r+N-1)!\,c^N}{(r-1)!\,\nu_1! \cdots \nu_n!\,(1+nc)^{r+N}}$$

$$= \sum_{t_n=M-r}^{\infty} \sum_{t_{n-1}=0}^{M-r-1} \sum_{t_{n-2}=0}^{t_{n-1}}$$

$$\cdots \sum_{t_1=0}^{t_2} \frac{(r+t_n-1)!\,c^{t_n}}{(r-1)!(t_n-t_{n-1})! \cdots (t_2-t_1)!\,t_1!(1+nc)^{r+t_n}}$$

$$= \sum_{t_n=M-r}^{\infty} \sum_{t_{n-1}=0}^{M-r-1} \frac{(r+t_n-1)!\,c^{t_n}(n-1)^{t_{n-1}}}{(r-1)!\,(t_n-t_{n-1})!\,t_{n-1}!\,(1+nc)^{r+t_n}}$$

$$= \sum_{u=0}^{\infty} \sum_{t_{n-1}=0}^{M-r-1} \frac{(M+u-1)!\,c^{M-r+u}(n-1)^{t_{n-1}}}{(r-1)!(M-r+u-t_{n-1})!\,t_{n-1}!\,(1+nc)^{M+u}}.$$

As before, all sums without indices are over all sequences $\nu_1, \cdots, \nu_n$ satisfying (6.1).

Since

$$\sum_{t_{n-1}=0}^{M-r-1} \frac{(n-1)^{t_{n-1}}}{t_{n-1}!\,(M-r+u-t_{n-1})!} = n^{M-r+u} \int_{1-1/n}^{1} \frac{t^{M-r-1}(1-t)^u}{(M-r-1)!\,u!}\,dt,$$

$$g(n) = \sum_{u=0}^{\infty} \frac{(M+u-1)!(nc)^{M-r+u}}{(M-r-1)!(r-1)!\,u!(1+nc)^{M+u}} \int_{1-1/n}^{1} t^{M-r-1}(1-t)^u\,dt$$

$$= \int_{1-1/n}^{1} \frac{(M-1)!(nc)^{M-r}}{(M-r-1)!(r-1)!} \frac{t^{M-r-1}}{(1+nc)^M} \cdot \left[\sum_{u=0}^{\infty} \frac{(M+u-1)!}{(M-1)!\,u!}\left(\frac{nc}{1+nc}\right)^u (1-t)^u\right] dt$$

$$= \int_{1-1/n}^{1} \frac{(M-1)!(nc)^{M-r}t^{M-r-1}}{(M-r-1)!(r-1)!(1+nct)^M}\,dt.$$

Writing $\tau = nct / (1 + nct)$, we have

$$g(n) = \frac{(M-1)!}{(r-1)!(M-r-1)!} \int_{\mathfrak{D}} \tau^{M-r-1}(1-\tau)^{r-1}\, d\tau,$$

$$\mathfrak{D} = \left[\frac{c(n-1)}{1+c(n-1)}, \frac{cn}{1+cn}\right].$$

From this we determine the expected value of n as a function of M, r, and c.

$$E(n) = \sum_{n=1}^{\infty} ng(n) = \sum_{n=1}^{\infty} g(n) + \sum_{n=2}^{\infty} g(n) + \cdots$$

$$= \frac{(M-1)!}{(r-1)!(M-r-1)!} \sum_{n=1}^{\infty} \int_{\mathcal{S}} \tau^{M-r-1}(1-\tau)^{r-1}\, d\tau,$$

$$\mathcal{S} = \left[\frac{c(n-1)}{1+c(n+1)}, 1\right].$$

8. Modified sequential procedure. Yet another sampling procedure yields an estimate of λ with relative error bounded by γ at confidence level α. This is the procedure in which we first obtain an observation a_r and then count the number of particles falling in a given number n_0 of non-overlapping areas, each of size ca_r, where n_0 is determined by the selected values of γ, α, c, and r. From the observations a_r and $\nu_1, \cdots, \nu_{n_0}$, we estimate λ by

$$\hat{l}^* = \frac{1-\gamma^2}{2\gamma a_r(1+n_0c)} \left[r + \sum_{i=1}^{n_0} \nu_i\right] \log \frac{1+\gamma}{1-\gamma}. \tag{8.1}$$

We state, with confidence coefficient α, that the relative error of $\hat{l}^*$ is bounded by γ. Consider the confidence coefficient with which we make this statement for any arbitrary value of n_0. The joint probability distribution of a_r and $\nu_1, \cdots, \nu_{n_0}$ is defined by the function ϕ given in (6.1), with no restrictions on the values of the ν_i and with $n = n_0$. Hence, the probability, before any observations are made, that

$$1 - \gamma \leqq \hat{}^* / \lambda \leqq 1 + \gamma$$

is given by

$$\sum_{\nu_1=0}^{\infty} \sum_{\nu_2=0}^{\infty} \cdots \sum_{\nu_{n_0}=0}^{\infty} \psi_\gamma^*\left(r + \sum_{i=1}^{n_0} \nu_i\right) \cdot h(\nu_1, \cdots, \nu_{n_0}, n_0),$$

and is independent of λ. For fixed values of r and c, this is a monotone increasing function of n_0. By a proper choice of n_0, the value of the confidence coefficient can be made as close to 1 as desired.

9. Estimate of bounded relative error for the variance of a normal distribution. In all of the sampling procedures yielding estimates of bounded relative error for the problem of particle counting, the function ψ_γ^* plays an important

part. This function also appears in the problem of estimates of bounded relative error for the variance of a normal distribution. Suppose we have N independent observations $x_1, \cdots, x_N$ on a chance variable X, where now X is normally distributed with unknown mean μ and variance σ^2. Our problem is to estimate σ^2, and we would like our estimate to be of bounded relative error.

If $w^2 = \sum_1^N (x_i - \bar{x})^2$, the distribution of w^2 / σ^2 is a chi-square distribution with $N - 1$ degrees of freedom. In this distribution, σ^2 appears as a scale parameter and is the only unknown parameter in the distribution. With the use of w^2, therefore, we can find an estimate of bounded relative error possessing specified optimal properties. For given values of γ and $n = N - 1$, the estimate $\hat{\sigma}^2$ which maximizes the confidence coefficient with which we state that

$$1 - \gamma \leqq \hat{\sigma}^2 / \sigma^2 \leqq 1 + \gamma$$

is given by

$$\hat{\sigma}^2 = 2\gamma w^2 \Big/ \left[n \log \frac{1 + \gamma}{1 - \gamma} \right],$$

$$P\{1 - \gamma \leqq \hat{\sigma}^2/\sigma^2 \leqq 1 + \gamma\} = \frac{1}{\Gamma(n/2)} \int_{\mathfrak{D}} 2^{-n/2} (\chi^2)^{(n-2)/2} e^{-\chi^2/2} \, d(\chi^2)$$

$$= \psi_\gamma^* \left(\frac{n}{2} \right),$$

$$\mathfrak{D} = \left[\frac{n(1 + \gamma)}{2\gamma} \log \frac{1 + \gamma}{1 - \gamma}, \frac{n(1 - \gamma)}{2\gamma} \log \frac{1 + \gamma}{1 - \gamma} \right].$$

Thus, making the substitution $n = 2M$, Table 1 gives us for $n = 4, 8, \cdots, 80$, values of the confidence coefficient with which we assert relation (8.1). As before, we can determine n so that $\psi_\gamma^*(n/2) = \alpha$ with α chosen in advance. For larger values of n, $\sqrt{2\chi_n^2} - \sqrt{2n - 1}$ is approximately normally distributed with zero mean and unit variance, and corresponding approximations can be made.

REFERENCES

[1] G. E. Albert and L. Nelson, "Contributions to the statistical theory of counter data," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 9–22.

[2] D. Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions*, John Wiley and Sons, New York, 1954.

[3] M. A. Girshick and L. J. Savage, "Bayes and minimax estimates for quadratic loss functions," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 13–22.

A CHARACTERIZATION OF SUFFICIENCY[1]

BY R. R. BAHADUR

The University of Chicago

Summary. The main conclusion of this paper can be described as follows. Consider a statistical decision problem in which certain structural conditions are satisfied, and let T be a statistic on the sample space. Then the class of decision functions which depend on the sample point only through T is essentially complete if and only if T is a sufficient statistic. The structural conditions in question are satisfied in many estimation problems.

1. Introduction. In a non-sequential decision problem, let $X = \{x\}$ be the sample space, $P = \{p\}$ the set of alternative probability distributions on X, $D = \{t\}$ the (terminal) decision space, and $L_p(t)$ the loss incurred in making the decision t when the (unknown) distribution on X is p. It is assumed that P is a dominated set of distributions (i.e., there exists a fixed σ-finite λ such that each p in P admits a probability density function with respect to λ), and that D is, or may be taken to be, a subset of k-dimensional euclidean space ($1 \leqq k \leqq \infty$).

For each decision function μ let the corresponding risk function be denoted by r_μ, that is, for each p in P, $r_\mu(p) =$ the expected value of L_p in using μ. Let $\mathfrak{D}$ be the class of all decision functions on X to D. Let $T(x)$ be a function on X (onto an arbitrary space Y of points y), and let $\mathfrak{D}_T$ be the class of all μ in $\mathfrak{D}$ which depend on x only through T. The class $\mathfrak{D}_T$ is said to be essentially complete if for each μ in $\mathfrak{D}$ there exists a ν in $\mathfrak{D}_T$ such that $r_\nu(p) \leqq r_\mu(p)$ for each p in P.

T is said to be a sufficient statistic for P if, for each set A of X and each value y of T, the conditional probability of A given $T(x) = y$ is the same for each p in P. It is well known (see, for example, [1], [2], [3]) that if T is sufficient for P, then $\mathfrak{D}_T$ is equivalent to $\mathfrak{D}$ in the sense that for each μ in $\mathfrak{D}$ there exists a ν in $\mathfrak{D}_T$ such that $r_\nu(p) = r_\mu(p)$ for each p in P.

It is shown in this paper that if the loss function L satisfies condition III of Section 2, and $\mathfrak{D}_T$ is essentially complete, then T must be sufficient foı P. Consequently, if III is satisfied, then, for any statistic T, any one of the statements "$\mathfrak{D}_T$ is essentially complete," "$\mathfrak{D}_T$ is equivalent to $\mathfrak{D}$," and "T is sufficient for P" implies the other two.

The following are some simple examples of decision problems in which condition III is satisfied.

EXAMPLE 1. Let P be a finite set, $P = \{p_1, p_2, \cdots, p_n\}$ say, let D be the set $\{1, 2, \cdots, n\}$, and let $L_i(j) = 0$ if $i = j$ and $=1$ if $i \neq j$ for $i, j = 1, 2, \cdots, n$.

Received July 26, 1954.

[1] Research supported by the Office of Naval Research under Contract No. N6onr-271, T.O. XI, Project No. 042-034.

In the following examples it is assumed that P is a parametric family of distributions, $P = \{p_\theta\}$ say, where θ is a real-valued parameter, and that the set Ω of all values of θ is an interval of the real line, say $\Omega = \{\theta : \theta_0 < \theta < \theta_1\}$, where $-\infty \leqq \theta_0 < \theta_1 \leqq \infty$.

EXAMPLE 2. Let $D = \Omega$, and let $L_\theta(t) = (t - \theta)^2$.

EXAMPLE 3. Let D be the set of all those points (r, s) of the plane for which $\theta_0 < r \leqq s < \theta_1$, and let each point $t = (r, s)$ of D correspond to the decision that $r \leqq \theta \leqq s$. Let $L_\theta(t) = h \cdot W_\theta(t) + k \cdot (s - r)$, where h and k are positive constants, and $W_\theta = 0$ if $r \leqq \theta \leqq s$ and $=1$ otherwise.

The preceding examples suggest that condition III is typical of problems of inference in which "nuisance parameters" are not involved. This is indeed the case.

2. Results. Let there be given: an abstract space X of points x, and a σ-field $\mathbf{S}$ of subsets of X; a set P of probability measures p on $\mathbf{S}$; a set D of points t, and a σ-field $\mathbf{D}$ of subsets of D; and a real-valued non-negative function $L_p(t)$ on $P \times D$ such that, for each p, L_p is a $\mathbf{D}$-measurable function of t. It is assumed that $(X, \mathbf{S})$, P, $(D, \mathbf{D})$, and L satisfy the following conditions.

CONDITION I. The set P of measures on $\mathbf{S}$ is a dominated set containing at least two measures.

It should perhaps be stated here that in the case when P contains only one measure, the conclusions of this paper hold entirely trivially.

CONDITION II. The decision space $(D, \mathbf{D})$ is of type $(R, \mathbf{R})$ ([3], Section 7), and D contains at least two points.

Let the closed interval $[0, 1]$ be denoted by I, and let $\mathbf{I}$ be the class of Borel sets of I. Let p and q be two different measures in P, and define

$$\alpha(u, t) = u \cdot L_p(t) + (1 - u) \cdot L_q(t) \tag{1}$$

for u in I and t in D, and

$$\beta(u) = \inf_{t \in D} \{\alpha(u, t)\} \tag{2}$$

for u in I. We suppose that there exists a function, τ say, on I into D, such that (i) τ is an $\mathbf{I}$-$\mathbf{D}$-measurable transformation, and (ii)

$$\alpha(u, \tau(u)) = \beta(u) \qquad \text{for each } u \text{ in } I. \tag{3}$$

In general, the function τ depends, of course, on the p and q under consideration.

An additional condition which we require is that the loss function L be quite sensitive, in a certain sense, to the difference between p and q. One condition of the type required is that the function τ of the preceding paragraph be uniquely determined and one to one. This condition is, however, unnecessarily strong, and it can be weakened, with advantage, as follows.

Let $c_1, c_2, \cdots$ be an enumeration of the rational points of I, excluding the end points 0 and 1. For each $i = 1, 2, \cdots$ define

$$\gamma_i(u) = c_i \cdot u + (1 - c_i) \cdot (1 - u), \qquad 0 < \gamma_i < 1, \tag{4}$$

and

$$\delta_i(u) = c_i \cdot u \,/\, \gamma_i(u), \qquad 0 \leqq \delta_i \leqq 1, \tag{5}$$

for u in I. We suppose that (iii) if u and v are any two points of I with $u \neq v$, then $\alpha(\delta_i(u), \tau(\delta_i(v))) > \beta(\delta_i(u))$ for at least one $i = 1, 2, \cdots$.

CONDITION III. Corresponding to any two measures p and q in P with $p \neq q$, there exists a function τ on I into D which satisfies (i), (ii), and (iii).

This condition is stated here in a form well adapted to our immediate purposes. It (or rather, its essential content) can be stated more simply as follows. Let Γ denote the zero-sum two-person game in which the spaces of pure strategies of players 1 and 2 are P and D respectively, and the payoff is L. For any p and q in P with $p \neq q$, let Γ_{pq} denote the subgame in which player 1 is restricted to the two pure strategies p and q and their mixtures. Then III is essentially the condition that each subgame Γ_{pq} is nontrivial and compact for player 2, in the following sense: there exists no t^* in D such that $L_p(t^*) \leqq L_p(t)$ and $L_q(t^*) \leqq L_q(t)$ for all t in D; and corresponding to each strategy of player 1, there exists a pure strategy of player 2 which minimizes the payoff. An amplification of this remark, in the form of a useful sufficient condition for III, is given in Section 3.

A decision function is a function on $\mathbf{D} \times X$, μ say, such that $\mu(C, x)$ is a probability measure on $\mathbf{D}$ for each x and an $\mathbf{S}$-measurable function of x for each C in $\mathbf{D}$ ([3], Section 7). For any decision function μ, the risk function r_μ is defined by

$$r_\mu(p) = \int_X \left\{ \int_D L_p(t)\, d\mu_x \right\} dp, \tag{6}$$

where, for each x, the expression in $\{\quad\}$ is the integral of L_p over D with respect to μ with x held fixed.

Let $\mathfrak{D}$ be the class of all decision functions. Let T be a function on X onto a set Y of points y, and let $\mathfrak{D}_T$ be the class of all ν in $\mathfrak{D}$ which are of the form $\nu = \nu(C, T(x))$.

THEOREM. *$\mathfrak{D}_T$ is an essentially complete subclass of $\mathfrak{D}$ if and only if T is a sufficient statistic for P.*

PROOF. Suppose first that T is sufficient for P, and consider a μ in $\mathfrak{D}$. It follows from Theorem 7.1 of [3], using condition II, that there exists a ν in $\mathfrak{D}_T$ such that

$$\int_X \mu(C, x)\, dp = \int_X \nu(C, T(x))\, dp \tag{7}$$

for all C in $\mathbf{D}$ and p in P. It follows from (7) and (6) that (for *any* loss function L) $r_\nu(p) = r_\mu(p)$ for each p in P. Since μ is arbitrary, we conclude that $\mathfrak{D}_T$ is essentially complete.

Suppose next that $\mathfrak{D}_T$ is essentially complete. Choose and fix p and q in P, with $p \neq q$, and define $\lambda(A) = p(A) + q(A)$ for A in $\mathbf{S}$. Since $p(A) \leqq \lambda(A)$, the Radon-Nikodým theorem yields the existence of an $\mathbf{S}$-measurable function,

$g(x)$ say, such that $0 \leqq g \leqq 1$, and

$$(8) \qquad p(A) = \int_A g \, d\lambda, \qquad q(A) = \int_A (1 - g) \, d\lambda$$

for all A in $\mathbf{S}$.

Let c_i be an arbitrary but fixed rational number such that $0 < c_i < 1$, and for any decision function μ define

$$(9) \qquad \bar{r}_\mu = c_i \cdot r_\mu(p) + (1 - c_i) \cdot r_\mu(q).$$

We see from (6) and (8) that the right side of (9) is equal to

$$c_i \int_X \left\{ \int_D L_p \, d\mu_x \right\} g \, d\lambda + (1 - c_i) \int_X \left\{ \int_D L_q \, d\mu_x \right\} (1 - g) \, d\lambda$$
$$= \int_X \left\{ \int_D [c_i g L_p + (1 - c_i)(1 - g) L_q] \, d\mu_x \right\} d\lambda.$$

Hence, for any μ,

$$(10) \qquad \bar{r}\mu = \int_X \gamma_i(g(x)) \left\{ \int_D \alpha(\delta_i(g(x)), t) \, d\mu_x \right\} d\lambda,$$

where α, γ_i, and δ_i are given by (1), (4), and (5).

Corresponding to the p and q under consideration, let τ be a function on I into D possessing properties (i), (ii) and (iii) of condition III. Let ξ be the non-randomized decision function which, for each x, assigns probability 1 to the decision $\tau(\delta_i(g(x)))$. It then follows from (2), (3), and (10) that

$$(11) \qquad \bar{r}_\xi = \inf_{\mu \epsilon \mathfrak{D}} \{\bar{r}_\mu\} = \int_X \gamma_i(g) \, \beta(\delta_i(g)) \, d\lambda.$$

It is convenient at this stage to consider explicitly the sample space of the values of T. Let $\mathbf{T}$ be the σ-field of all sets $B \subset Y$ such that $T^{-1}(B)$ is in $\mathbf{S}$, and for any measure m on $\mathbf{S}$ denote the induced measure on $\mathbf{T}$ by m^*, that is, $m^*(B) = m(T^{-1}(B))$. We have been regarding $\mathfrak{D}_T$ as a class of decision functions on $(X, \mathbf{S})$, but $\mathfrak{D}_T$ can also be regarded as the class, $\mathfrak{D}^*$ say, of all decision functions on $(Y, \mathbf{T})$. In particular, $\nu(C, T(x)) \leftrightarrow \nu(C, y)$ is a one to one correspondence between $\mathfrak{D}_T$ and $\mathfrak{D}^*$ such that, for any probability measure m on $\mathbf{S}$,

$$r_\nu(m) \equiv \int_X \left\{ \int_D L_m(t) \, d\nu_{T(x)} \right\} dm = \int_Y \left\{ \int_D L_m(t) \, d\nu_y \right\} dm^* = r_\nu(m^*).$$

By replacing X, $\mathbf{S}$, p, q, and $\mathfrak{D}$ by Y, $\mathbf{T}$, p^*, q^*, and $\mathfrak{D}^*$ in the argument leading to (11), and then rephrasing the outcome according to the correspondence just stated, we obtain the following result. There exists a non-randomized decision function in $\mathfrak{D}_T$, η say, such that

$$(12) \qquad \bar{r}_\eta = \inf_{\nu \epsilon \mathfrak{D}_T} \{\bar{r}_\nu\},$$

and such that, for each x, η assigns probability 1 to the decision $\tau(\delta_i(f(x)))$, where

$$f(x) \equiv h(T(x)) \tag{13}$$

and $h(y)$ is a T-measurable function on Y (depending only on p^* and q^*) such that $0 \leqq h \leqq 1$.

Consider the decision function ξ defined above. Since $\mathfrak{D}_T$ is essentially complete, there exists a ν in $\mathfrak{D}_T$ such that $r_\nu \leqq r_\xi$ for each measure in P; in particular, $r_\nu(p) \leqq r_\xi(p)$ and $r_\nu(q) \leqq r_\xi(q)$. Hence, $\bar{r}_\nu \leqq \bar{r}_\xi$ by (9). It now follows from (11) and (12) that $\bar{r}_\eta = \bar{r}_\xi$. Using (10) to evaluate $\bar{r}_\eta$ (cf. the definition of η), it is easily seen from (11) that this last equality is

$$\int_X \gamma_i(g)\cdot\alpha(\delta_i(g), \tau(\delta_i(f)))\, d\lambda = \int_X \gamma_i(g)\cdot\beta(\delta_i(g))\, d\lambda. \tag{14}$$

Now $0 \leqq \beta(u) \leqq \inf_t \{\max [L_p(t), L_q(t)]\} < \infty$ by (1) and (2), and $0 < \gamma_i(u) < 1$ by (4), so that $\gamma_i(u)\cdot\beta(u)$ is a bounded function of u. Since λ is a finite measure, it follows that the right side of (14) is finite. Hence, writing

$$\varphi_i(x) \equiv \alpha(\delta_i(g(x)), \tau(\delta_i(f(x)))) - \beta(\delta_i(g(x))), \tag{15}$$

(14) is equivalent to

$$\int_X \gamma_i(g)\cdot\varphi_i\, d\lambda = 0. \tag{16}$$

Since $\varphi_i \geqq 0$ for each x by (2) and (15), and $\gamma_i(g) > 0$ for each x by (4), (16) implies that there exists an S-measurable set, N_i say, such that $\lambda(N_i) = 0$, and $\varphi_i(x) = 0$ for each x in $X - N_i$.

Since in the preceding argument c_i is arbitrary, it follows that there exists an S-measurable set $N \quad (= \bigcup_i N_i)$ such that $\lambda(N) = 0$ and such that for each x in $X - N$ we have $\varphi_i(x) = 0$ for $i = 1, 2, \cdots$. Hence by the definition (15) of the sequence $\varphi_1, \varphi_2, \cdots$, condition III(iii), and (13), we have $g(x) = h(T(x))$ on $X - N$. Write $\psi_p(y) = h(y)$ and $\psi_q(y) = 1 - h(y)$. Then $\psi_p(T(x))$ and $\psi_q(T(x))$ are non-negative S-measurable functions of x, such that $dp = \psi_p T\, d\lambda$ and $dq = \psi_q T\, d\lambda$ on S, by (8) and the equivalence of g and hT which we have just established. Hence, by the factorization theorem for sufficient statistics, T is a sufficient statistic for the set $P_0 = \{p, q\}$.

Since in the preceding argument p and q are arbitrary, we have shown that T is pairwise sufficient for P in the sense of [1]. It now follows from Theorem 2 of [1], using condition I, that T is sufficient for P. This completes the proof of the theorem.

The following comments concerning the condition III are relevant to the theorem of this section. (a) If the given loss function $L_p(t)$ satisfies III, then so does any loss function of the form $k(p)\cdot L_p(t)$ where $0 < k(p) < \infty$ for each p in P. This is as it should be, since essential completeness of a class $\mathfrak{D}_T$ is invariant under such modifications of the loss function. (b) For each p in P, let

D_p be the set of all points in D which minimize $L_p(t)$. It is easily seen that III implies that $\{D_p\}$ is a family of nonempty and disjoint subsets of D. Consequently, III cannot be satisfied in reasonable formulations of problems such as testing hypotheses (except when the hypothesis and the alternative are both simple), or estimating a parameter θ such that more than one p in P has the same value of θ. This, again, is not unexpected, since in such problems the useful concept is that of "sufficiency for the relevant parameter" rather than the unrestricted sufficiency with which the theorem is concerned. (c) Condition III is not, however, necessary to the theorem. It can be shown by examples that if III is not satisfied, the theorem may or may not hold.

Let $\mathfrak{A}$ be the class of all admissible decision functions.

COROLLARY. *Suppose that $\mathfrak{A}$ is a complete class. Then a statistic T is sufficient for P if and only if for each μ in $\mathfrak{A}$ there exists a ν in $\mathfrak{D}_T$ such that $r_\nu(p) = r_\mu(p)$ for each p in P.*

PROOF. It follows from the definition of admissibility that if $\mathfrak{A}$ is complete, then a class $\mathfrak{D}_T$ possesses the property stated if and only if $\mathfrak{D}_T$ is essentially complete, and the theorem applies.

3. Condition IV. Let p and q be measures in P, and consider the function $F_{pq}(t) = [L_p(t), L_q(t)]$ which maps D into the plane. Let J_{pq} be the range of F_{qp}. Let us say that a point (r, s) in J_{pq} is admissible if there exists no (r^*, s^*) in J_{pq} such that $r^* \leqq r, \quad s^* \leqq s$, and $r^* + s^* < r + s$. Let K_{pq} be the set of all admissible points of J_{pq}. Let us also say that K_{pq} is a complete subset of J_{pq} if for each (r, s) in J_{pq}, there exists an (r^*, s^*) in K_{pq} such that $r^* \leqq r, \quad s^* \leqq s$. The terms "admissible" and "complete" are borrowed from statistical decision theory, but as used here they refer not to the statistical decision problem, nor even to the game Γ, but only to the subgame Γ_{pq} (cf. Section 2).

It is well known (and easily shown by examples) that K_{pq} can be the empty set, and that even if K_{pq} is nonempty it need not be complete.

CONDITION IV. For any two measures p and q in P with $p \neq q$, (a) K_{pq} contains at least two points, (b) K_{pq} is a closed and bounded subset of the plane, (c) K_{pq} is a complete subset of J_{pq}, and (d) for each (r, s) in K_{pq} there exists only one point t in D such that $F_{pq}(t) = (r, s)$.

It can be shown that II and IV imply III. An outline of the proof follows.

Consider fixed p and q in P with $p \neq q$. Parts (a), (b) and (c) of IV assure the existence of a τ which satisfies parts (ii) and (iii) of III. The additional conditions IV(d) and II (together with the fact that the inverse of a 1–1 Borel measurable function is Borel measurable) assure that τ also satisfies III(i). The construction and detailed verification of τ is, however, rather lengthy, and it seems best to omit it. Note that since the conditions IV(d) and II are used only to assure that τ is measurable, they are superfluous in applications where the measurability of τ is not in doubt.

The conditions II and IV are satisfied in each of the examples of Section 1. Indeed, they are satisfied in Example 1 with any L such that $L_i(j) > L_i(i)$

whenever $i \neq j$; in Example 2 with $L_\theta(t) = a(\theta) \cdot |t - \theta|^{b(\theta)}$, where $0 < a(\theta) < \infty$ and $0 < b(\theta) < \infty$; and in Example 3 with $L_\theta(r, s) = A_\theta(r, s) + B_\theta(s - r)$ where $A_\theta = 0$ if $r \leqq \theta \leqq s$ and > 0 otherwise, and $B_\theta(z)$ is a strictly increasing function of $z \geqq 0$ with $B_\theta(0) = 0$.

4. A theorem of Elfving. In this section we suppose that there are given $(X, \mathbf{S})$, P, and $(D, \mathbf{D})$, as before, but that the loss function L is not specified. The class $\mathfrak{D}_T$ is said to be uniformly essentially complete if, for every loss function L, $\mathfrak{D}_T$ is essentially complete. The concept of uniform essential completeness is due to Elfving [4]. In his paper, he showed (using a notation and terminology which differs slightly from the present one) that if each of the sets X, P, and D is finite, then $\mathfrak{D}_T$ is uniformly essentially complete if and only if T is sufficient for P. We shall show that this result is valid provided only that P and D satisfy conditions I and II.

In view of the first part of the proof of the theorem in Section 2, we need only show that if $\mathfrak{D}_T$ is uniformly essentially complete, then T is sufficient. Let p and q be two measures in P, $p \neq q$, and let $P_0 = \{p, q\}$. It is easily seen that the hypothesis implies that $\mathfrak{D}_T$ is uniformly essentially complete for $(X, \mathbf{S})$, P_0, and $(D, \mathbf{D})$. Let r and s be points of D, $r \neq s$, and define

$$L_p(t) = \begin{cases} 0 & \text{if } t = r \\ 1 & \text{if } t = s \\ 2 & \text{otherwise,} \end{cases}$$

$$L_q(t) = \begin{cases} 1 & \text{if } t = r \\ 0 & \text{if } t = s \\ 2 & \text{otherwise.} \end{cases}$$

Then, as is easily seen, I, II, and III are satisfied in the problem $(X, \mathbf{S})$, P_0, $(D, \mathbf{D})$, and L. Since $\mathfrak{D}_T$ is essentially complete for this problem in particular, it follows from the theorem of Section 2 that T is sufficient for $P_0 = \{p, q\}$. Since p and q are arbitrary, it follows from Theorem 2 of [1] that T is sufficient for P, as was to be shown.

Let us say that $\mathfrak{D}_T$ is uniformly equivalent to $\mathfrak{D}$ if, for every loss function L, $\mathfrak{D}_T$ is equivalent to $\mathfrak{D}$ in the sense of Section 1. Let us also say $\mathfrak{D}_T$ is strongly equivalent to $\mathfrak{D}$ if corresponding to each μ in $\mathfrak{D}$ there exists a ν in $\mathfrak{D}_T$ such that $\int_X \nu(C, T(x))\, dp = \int_X \mu(C, x)\, dp$ for all C in $\mathbf{D}$ and p in P. Now, it is shown in [3] that sufficiency implies strong equivalence. The facts that strong equivalence implies uniform equivalence implies uniform essential completeness are evident, and we have just seen that uniform essential completeness implies sufficiency. Consequently, the concepts of sufficiency, strong equivalence, uniform equivalence, and uniform essential completeness afford equivalent comparisons of $\mathfrak{D}_T$ and $\mathfrak{D}$, at least when I and II are satisfied.

The conclusion just stated could be regarded as a strong result in the comparison of experiments in the special case when one of the two experiments being compared is a contraction of the other (cf. [5]). By so regarding it, it follows, in

particular, that (at least in the case when $(X, \mathbf{S})$ is of type $(R, \mathbf{R})$ and P is dominated) a statistic $y = T(x)$ is sufficient for x in Blackwell's sense [5] if and only if y is sufficient for x in the classical sense, that is to say, T is sufficient for P.

REFERENCES

[1] P. R. HALMOS AND L. J. SAVAGE, "Application of the Radon-Nikodym theorem to the theory of sufficient statistics," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 225–241.

[2] DAVID BLACKWELL AND M. A. GIRSHICK, *Theory of Games and Statistical Decisions*, John Wiley and Sons, New York, 1954.

[3] R. R. BAHADUR, "Sufficiency and statistical decision functions," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 423–462.

[4] G. ELFVING, "Sufficiency and completeness in decision function theory," *Ann. Acad. Sci. Fennicae*, Ser. A, I. Math.-Phys., No. 135 (1952).

[5] DAVID BLACKWELL, "Equivalent comparisons of experiments," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 265–272.

[6] A. WALD, *Statistical Decision Functions*, John Wiley and Sons, New York, 1950.

DISTRIBUTION OF THE MAXIMUM OF THE ARITHMETIC MEAN OF CORRELATED RANDOM VARIABLES

By John Gurland

Iowa State College

1. Summary. The initial distribution considered here is obtained from a multivariate analogue of the Pearson Type III distribution, and the value of the correlation is taken to be non-negative. There is obtained here the distribution of the maximum in samples of fixed size n from a random variable which is the arithmetic mean of k such correlated random variables. This distribution is obtained for large values of n and for large values of k. The appropriate expressions for the mode and scale parameters are also given.

2. Introduction. The mathematical model presented here is applicable when the aforementioned samples of size n consist of independent observations from a fixed population. The distribution function for this population, given by (4) below, is obtained by regarding each of the k correlated random variables as having the same fixed Pearson Type III distribution, and as having the same correlation with each of the remaining variables. Such a model may be relevant, for instance, in considering the tensile strength of a substance; in this case n is roughly proportional to the number of flaws. Another possible field of application is the investigation of maximum coincident loads in electrical engineering problems; but in this case there are difficulties involved in determining the appropriate value of n, and in assuming that the n observations of the sample are independent. The above mathematical model could be extended to cover such situations; however, the present article is confined to the theoretical problem stated in the summary.

3. Distribution of the arithmetic mean of correlated Pearson Type III variables. In many problems the primary distribution is skew and is well fitted by a Pearson Type III distribution

$$p_{z_1}(x) = \frac{e^{-x/\theta}x^{\lambda-1}}{\theta\Gamma(\lambda)}, \qquad x > 0. \tag{1}$$

Before considering the problem concerning the maximum value of the mean, it is first required to ascertain the distribution of $(1/k)\sum_1^k Z_i$, where each Z_i has the probability density (1) and there is a constant correlation ρ between every pair of Z's. The assumption of constant correlation is indeed restrictive, but it is most convenient to have a single parameter which measures the over-all correlation of the population. It is possible to consider more general models by the same methods described below, but the present article considers only the case of constant correlation.

Received September 21, 1953, revised August 17, 1954.

The characteristic function of a multivariate analogue of (1) is obtained by a device, similar to that used by Krishnamoorthy and Parthasarathy [1], as follows. Consider the multivariate normal probability density

$$p_X(x) = \frac{|\Omega|^{1/2}}{(\sigma\sqrt{2\pi})^k} \exp\left(-\frac{x\Omega x'}{2\sigma^2}\right),$$

where $X = (X_1, X_2, \cdots, X_k)$ and Ω is a positive definite matrix. The characteristic function of $Y = (X_1^2, X_2^2, \cdots, X_k^2)$ is

$$\phi_Y(t) = \frac{|\Omega|^{1/2}}{|\Omega - 2i\sigma^2 T|^{1/2}},$$

where T is the diagonal matrix

$$T = \begin{bmatrix} t_1 & 0 & 0 & \cdots & 0 \\ 0 & t_2 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & t_k \end{bmatrix}$$

Then $[\phi_Y(t)]^m$ is the characteristic function of a natural multivariate analogue of the χ^2 distribution. Now

$$[\phi_Y(t)]^m = |I - 2i\sigma^2 T\Omega^{-1}|^{-m/2}$$

where I is the unit $n \times n$ matrix. Take $\frac{1}{2}m = \lambda$ and $2\sigma^2 = \theta$ to obtain

$$\phi_Z(t) = |I - i\theta T\Omega^{-1}|^{-\lambda}. \tag{2}$$

This is the characteristic function of $Z = (Z_1, Z_2, \cdots, Z_k)$, where each Z_i has the probability density (1). Note that in (2) any positive definite covariance matrix Ω is permissible. The special case of interest here is

$$\Omega^{-1} = \begin{bmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho & \rho & \rho & \cdots & 1 \end{bmatrix}$$

The characteristic function of $(1/k)\sum_1^k Z_i$ is then

$$f(t) = \begin{vmatrix} 1 - it\theta/k & -it\theta\rho/k & \cdots & -it\theta\rho/k \\ -it\theta\rho/k & 1 - it\theta/k & \cdots & -it\theta\rho/k \\ \cdots & \cdots & \cdots & \cdots \\ -it\theta\rho/k & -it\theta\rho/k & \cdots & 1 - it\theta/k \end{vmatrix}^{-\lambda} \tag{3}$$

$$= [1 - (it\theta/k)(1 - \rho)]^{-\lambda(k-1)}[1 - (it\theta/k)(1 - \rho + \rho k)]^{-\lambda}.$$

On referring to tables of Fourier integrals [2], one finds that the probability density is

$$(4)\qquad p(x) = \frac{1}{c}\, x^{k\lambda-1} \exp\left[\frac{-kx}{\theta(1-\rho)}\right] {}_1F_1\left(\lambda, k\lambda, \frac{\rho k^2 x}{\theta(1-\rho)(1-\rho+\rho k)}\right)$$

where

$$(5)\qquad c = (\theta/k)^{\lambda k}\,(1-\rho)^{\lambda(k-1)}\,(1-\rho+\rho k)^{\lambda}\,\Gamma(\lambda k)$$

and where ${}_1F_1(w, v, z)$ is the confluent hypergeometric function which is given by

$${}_1F_1(w, v, z) = 1 + \frac{w}{v}\frac{z}{1} + \frac{w(w+1)}{v(v+1)}\frac{z^2}{2!} + \frac{w(w+1)(w+2)}{v(v+1)(v+2)}\frac{z^3}{3!} + \cdots.$$

For $\rho = 0$ the density given by (4) reduces to that of an arithmetic mean of k independent random variables, each of which has the density given by (1).

4. Asymptotic solution of a certain transcendental equation. Before obtaining the large-sample distribution of the maximum value in samples from (4), it is necessary to consider the solution of the equation

$$(6)\qquad x - A \log x + B = C$$

where A, B, and C are constants and the value of C is large. If $A < 0$ there is one solution but if $A > 0$ there are two solutions, as is easily seen by considering the intersection of $y = x + B$ and $y = A \log x + C$, and noting that $C \to \infty$. However the only solution required here is the one which increases indefinitely as C increases indefinitely.

Let $r(x) = x - A \log x + B - C$. Now $dr/dx = 1 - A/x > 0$ for x large, which shows that the function is monotonically increasing for x sufficiently large. As a first approximation to a solution, try

$$(7)\qquad x_1 = C - B + A \log C.$$

This makes

$$(8)\qquad r(x_1) = -A \log\left\{1 - \frac{B - A \log C}{C}\right\} \approx \frac{A\,(B - A \log C)}{C}$$

for C large. This approaches zero as $C \to \infty$. It can, in fact, be shown that the solution satisfies

$$(9)\qquad x = C - B + A \log C + o(1/C^{1-\epsilon})$$

where ϵ is an arbitrarily small number such that $0 < \epsilon < 1$. This is easily seen[1] by noting that for C large enough,

$$r(C - B + A \log C - 1/C^{1-\epsilon/2}) < 0,$$
$$r(C - B + A \log C + 1/C^{1-\epsilon/2}) > 0.$$

[1] The author is grateful to the referee for simplifying this proof.

If B as well as C is permitted to increase indefinitely, the asymptotic solution (7) remains valid, provided $B/C \to 0$. This requirement is obvious on referring to (8). However, the order of the approximation indicated in (9) may not be valid for this case. A sufficient condition for the validity of (9) is $B = O(\log C)$.

5. Large sample distribution of the maximum for the case $\rho > 0$. The main problem now is to find the distribution function of the maximum in samples of size n from the population characterized by the density (4). In this section the distribution will be obtained first for large values of n, then for large values of k and n. The two cases $\rho > 0$ and $\rho = 0$ will also be treated separately.

Let $F(x) = \int_0^x p(t)\,dt$ where $p(t)$ is the density (4). Then $[F(x)]^n$ is the distribution function of the maximum in samples of size n. The probability density of the maximum is

$$g(x) = nF^{n-1}(x)\,p(x). \tag{10}$$

Apply the transformation

$$y_n = n[1 - F(x)]. \tag{11}$$

Then, if X is the random variable with density (10), and Y_n is the random variable defined by (11), Y_n has the probability density function

$$p_{Y_n}(x) = (1 - x/n)^{n-1},$$

and it will be noted that $\lim_{n\to\infty} p_{Y_n}(x) = e^{-x}$, for $x > 0$. In solving (11) for x as a function of y_n, it will be shown below that $X = -\alpha \log Y_n + \beta$ for large values of n (and hence large values of x). Consequently, by a limit theorem of Mann and Wald [3], the limiting distribution function of $(X - \beta)/\alpha$ is the distribution function $\exp(-e^{-x})$.

We now proceed to solve (11) for x. The equation may be written

$$y_n = \frac{n}{c}\int_x^\infty t^{\lambda k-1}\exp\left[\frac{-kt}{\theta(1-\rho)}\right] {}_1F_1(\lambda, k\lambda, \delta t)\,dt \tag{12}$$

where $\delta = \rho k^2/\theta(1-\rho)(1-\rho+\rho k)$, and c is given by (5).

Since the distribution will be considered for large values of n it will suffice to find the solution of (12) which is valid for large values of x (cf. Fisher and Tippett [4], Cramér [5]). In evaluating the integral in (12), the following properties of the conflue t hypergeometric function will be required.

$$\frac{d}{dx}\,{}_1F_1(w, v, x) = \frac{w}{v}\;{}_1F_1(w+1,\; v+1,\; x); \tag{13}$$

$${}_1F_1(w, v, x) = \frac{e^x\Gamma(v)}{x^{v-w}\Gamma(w)}\left[1 + O\left(\frac{1}{x}\right)\right], \qquad x \to \infty. \tag{14}$$

Formula (13) follows from the definition of the confluent hypergeometric function (cf. [6]). A more general asymptotic expression than (14) is established by Barnes [7].

Integrate (12) by parts and apply (13) to obtain $cy_n/n = \alpha_1 + \alpha_2 + \alpha_3$, where

$$\alpha_1 = \frac{\theta(1-\rho)}{k} x^{k\lambda-1} \exp\left[\frac{-kx}{\theta(1-\rho)}\right] {}_1F_1(\lambda, k\lambda, \delta x),$$

$$\alpha_2 = \frac{\theta(1-\rho)(k\lambda-1)}{k} \int_x^\infty t^{\lambda k-2} \exp\left[\frac{-kt}{\theta(1-\rho)}\right] {}_1F_1(\lambda, k\lambda, \delta t)\, dt,$$

$$\alpha_3 = \frac{\delta\theta(1-\rho)}{k^2} \int_x^\infty t^{\lambda k-1} \exp\left[\frac{-kt}{\theta(1-\rho)}\right] {}_1F_1(\lambda+1, k\lambda+1, \delta t)\, dt.$$

Now define $\alpha = \theta(1-\rho+\rho k)/k$. In virtue of (14), and for $\rho > 0$, these expressions reduce to

$$\alpha_1 = \frac{\theta(1-\rho)}{k} \frac{x^{\lambda-1}}{\delta^{\lambda(k-1)}} \frac{\Gamma(k\lambda)}{\Gamma(\lambda)} e^{-x/\alpha}\left[1 + O\left(\frac{1}{\delta x}\right)\right],$$

$$\alpha_2 = \frac{\theta(1-\rho)\alpha}{k} \frac{k\lambda-1}{\delta^{\lambda(k-1)}} x^{\lambda-2} \frac{\Gamma(k\lambda)}{\Gamma(\lambda)} e^{-x/\alpha}\left[1 + O\left(\frac{1}{\delta x}\right)\right],$$

$$\alpha_3 = \frac{\theta(1-\rho)\alpha}{k} \frac{x^{\lambda-1}}{\delta^{\lambda(k-1)-1}} \frac{\Gamma(k\lambda)}{\Gamma(\lambda)} e^{-x/\alpha}\left[1 + O\left(\frac{1}{\delta x}\right)\right].$$

As a result of the above simplification it is now possible to write (12) in the form

$$\frac{cy_n}{n} = \frac{\theta(1-\rho)}{k} \frac{(1+\delta\alpha)x^{\lambda-1}}{\delta^{\lambda(k-1)}} \frac{\Gamma(k\lambda)}{\Gamma(\lambda)} e^{-x/\alpha}\left[1 + O\left(\frac{1}{\delta x}\right)\right].$$

A further simplification is afforded by referring to the representation of c in (5). Thus

$$\text{(15)} \qquad \begin{aligned} \frac{c_1 y_n}{n} &= x^{\lambda-1} e^{-x/\alpha}\left[1 + O\left(\frac{1}{\delta x}\right)\right], \\ c_1 &= \frac{\alpha^{\lambda k-1}}{1+\alpha}(1-\rho)^{\lambda(k-1)-1}(1-\rho+\rho k)^{\lambda(1-k)-1}\Gamma(\lambda). \end{aligned}$$

It now remains to solve (15) as a function of y_n. If the terms involving $O(1/\delta x)$ are neglected, and we take logarithms of the remaining terms, we obtain

$$\text{(16)} \qquad x - \alpha(\lambda-1)\log x + \alpha(\log y_n + \log c_1) = \alpha \log n.$$

Define

$$\text{(17)} \quad A = \alpha(\lambda-1), \qquad B = \alpha(\log y_n + \log c_1), \qquad C = \alpha \log n.$$

If k is small and n is large, then the solution (9) is applicable. Hence

$$\text{(18)} \qquad \begin{aligned} x &= -\alpha \log y_n + \beta + o\,(1/[\alpha \log n]^{1-\epsilon}), \\ \beta &= \alpha\,[\log(n/c_1) + (\lambda-1)\log(\alpha \log n)]. \end{aligned}$$

Thus, the limiting distribution of $(X-\beta)/\alpha$ is $\exp(-e^{-x})$, and α and β represent respectively the scale parameter and the mode for large values of n.

If k as well as n is allowed to increase indefinitely, then in order to apply the results of Section 4, the expressions for B and C in (17) must be altered. In fact, equation (16) may be written as

$$x - \alpha(\lambda - 1) \log x + \alpha \log y_n + \alpha \log \{\theta^{\lambda k-1}(1-\rho)^{\lambda(k-1)-1}(1-\rho+\rho k)^{\lambda}\delta^{\lambda(k-1)}\Gamma(\lambda)\} = \alpha \log [k^{\lambda k-2}\{k + \delta\theta(1-\rho+\rho k)\}] + \alpha \log n.$$

Now if we define

$$B = \alpha \log y_n + \alpha \log \{\theta^{\lambda k-1}(1-\rho)^{\lambda(k-1)-1}(1-\rho+\rho k)^{\lambda}\delta^{\lambda(k-1)}\Gamma(\lambda)\},$$
$$C = \alpha \log n + \alpha \log [k^{\lambda k-2}\{k + \delta\theta(1-\rho+\rho k)\}],$$

it follows, since δ is of the order of magnitude of k, that

$$B \approx \alpha\lambda k \log k, \qquad C \approx \alpha\lambda k \log k + \alpha \log n.$$

Hence, if $k \log k = o(\log n)$, it follows that $\lim_{k\to\infty, n\to\infty} B/C = 0$, in which case (9) is applicable. This gives $x = -\alpha \log y_n + \beta$, with $\alpha = \theta(1-\rho+\rho k)/k$ as in (16) and

(19) $\quad \beta = \alpha \log (n/c_1) + \alpha(\lambda - 1) \log \{\alpha \log [n(k\lambda)^{k\lambda-2}(k + \delta\theta(1-\rho+\rho k)]\}.$

6. Large sample distribution of the maximum for the case $\rho = 0$. The distribution of the maximum for the case $\rho = 0$ must be considered separately because δ now reduces to zero, and the approximations employed in the evaluation of α_1, α_2, and α_3 are no longer valid for large values of x. With $\rho = 0$ the density in (4) becomes

$$p(x) = \frac{1}{c} x^{k\lambda-1} e^{-kx/\theta}, \qquad x > 0$$

where now $c = (\theta/k)^{\lambda k}\Gamma(\lambda k)$. The same method as that described above is applied to

$$y_n = \frac{n}{c}\int_x^\infty t^{\lambda k-1} e^{-kt/\theta}\, dt.$$

For large values of x this yields

$$cy_n/n = (\theta/k)e^{-kx/\theta}x^{\lambda k-1}[1 + O(1/x)].$$

As before, neglect the terms involving $O(1/x)$ and take logarithms of the remaining terms to obtain

(20) $\quad x - \alpha(k\lambda - 1) \log x + \alpha[\log y_n + \log \{\alpha^{k\lambda-1}\Gamma(\lambda k)\}] = \alpha \log n$

where now $\alpha = \theta/k$. If k is small and n is large the solution becomes again $x = -\alpha \log y_n + \beta$, where now

(21) $\quad \beta = \alpha \log \dfrac{n}{\alpha^{\lambda k-1}\Gamma(\lambda k)} + \alpha(k\lambda - 1) \log (\alpha \log n).$

If k as well as n is allowed to increase indefinitely, then, as before, the expressions for B and C must be altered to apply the results of Section 4. Rewrite equation (20) as

$$x - \alpha(k\lambda - 1)\log x + \alpha \log y_n = \alpha \log n - \alpha \log [\alpha^{\lambda k-1}\Gamma(\lambda k)].$$

If we define

$$B = \alpha \log y_n, \qquad C = \log \frac{n}{\alpha^{\lambda k-1}\Gamma(\lambda k)},$$

then $\lim_{k\to\infty, n\to\infty} B/C = 0$ if $\alpha^{\lambda k-1}\Gamma(\lambda k) = o(n)$. We obtain, as before,

$$x = -\alpha \log y_n + \beta,$$

where now

$$\beta = \alpha \log \frac{n}{\alpha^{\lambda k-1}\Gamma(k\lambda)} + \alpha(k\lambda - 1)\log\left\{\alpha \log \frac{n}{\alpha^{\lambda k-1}\Gamma(k\lambda)}\right\}. \tag{22}$$

6. Conclusion. A few remarks are in order regarding the results obtained in this paper. Firstly, the cases in which both k and n are allowed to increase indefinitely require that k should not increase too rapidly relative to n, if the results obtained are to remain valid. Further, the order of approximation for these cases need not be the same as for the cases in which only n is allowed to increase indefinitely.

Secondly, some extensions of the results obtained here require further research. For instance, as has already been stated, the correlation need not be the same for all pairs of Z_i's. Further, the initial distribution from which samples of size n are taken need not be fixed but might change during the course of the sampling. This might further be complicated by the fact that the observations in the sample could be correlated.

7. Acknowledgments. The author has had many stimulating discussions with Dr. Emil Jebe and Mr. Landy Altman, who first brought this general problem to his attention. He has also had the benefit of some illuminating comments by Dr. Ray Mickey and Dr. Harry Goheen.

REFERENCES

[1] A. S. Krishnamoorthy and M. Parthasarathy, "A multivariate gamma-type distribution", *Ann. of Math. Stat.*, Vol. 22 (1951), pp. 549–557.

[2] G. A. Campbell and R. M. Foster, "Fourier integrals for practical applications", *Bell Telephone Technical Publications*, 1942.

[3] H. B. Mann and A. Wald, "On stochastic limit and order relationships", *Ann. of Math. Stat.*, Vol. 14 (1943), pp. 217–226.

[4] R. A. Fisher and L. H. C. Tippett, "Limiting forms of the frequency distribution of the largest or smallest member of a sample", *Proc. Cambridge Philos. Soc.*, Vol. 24 (1928), pp. 180–190.

[5] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[6] *Tables of the Confluent Hypergeometric Function $F(n/2, \frac{1}{2}, x)$ and Related Functions*, National Bureau of Standards Applied Mathematics Series, No. 3, U. S. Government Printing Office, Washington, D. C.

[7] E. W. Barnes, "On functions defined by simple types of hypergeometric series", *Trans. Cambridge Philos. Soc.*, Vol. 20 (1905), pp. 253–280.

SIGNIFICANCE PROBABILITIES OF THE WILCOXON TEST

By Evelyn Fix[1] and J. L. Hodges, Jr.

University of California, Berkeley

0. Summary. Tables are presented from which exact values of the Wilcoxon distribution may be obtained when the smaller sample size m does not exceed 12. The Edgeworth approximation to terms of order $1/m^2$ is given and its accuracy investigated.

1. Introduction. We are interested in the problem of obtaining significance probabilities for the Wilcoxon unpaired two-sample test [1], [2]. Let $m \leqq n$ be positive integers, and let $R_1 < R_2 < \cdots < R_m$ represent a random sample of size m drawn without replacement from the first $m + n$ positive integers. Let $S_i = R_i - i$ and $U = S_1 + S_2 + \cdots + S_m$, and let $\pi(u, m, n)$ denote the distribution function of U. It is the values of the function π that are required in the Wilcoxon test. [Wilcoxon actually considered $W = R_1 + \cdots + R_m = U + \frac{1}{2}m(m + 1)$].

Mann and Whitney [2] have tabled[2] π to $3D$ for $n \leqq 8$, and have shown that π, suitably normalized, tends to the normal as $m, n \to \infty$. White [3] has tabled the largest value of u for which $\pi(u, m, n) \leqq 0.005$, 0.025 for $m + n \leqq 30$. Auble [4] has published a similar table for $m, n \leqq 20$, and significance levels 0.001, 0.005, 0.01, 0.02, 0.025, 0.04, 0.05, 0.1. These tables and the normal approximation serve most ordinary needs in hypothesis testing. For some purposes (such as relative efficiency studies, in which it is the relative error that matters) the normal approximation is not sufficiently precise, and the restriction $n \leqq 8$ of the Mann-Whitney table is confining. The White and Auble tables give significance probabilities in most cases with even less accuracy than the normal approximation. [[3] contains several errors of one, apparently due to rounding.]

The connection of π with a partition function is well known. If $A(u, m, n)$ denotes the number of ways (without regard to order) in which it is possible to choose exactly m nonnegative integral summands, none greater than n, whose sum does not exceed u [or, equivalently, the number of ways in which it is possible to choose exactly m positive distinct integral summands, none greater than $m + n$, whose sum does not exceed $u + \frac{1}{2}m(m + 1)$], then

$$\pi(u, m, n) = A(u, m, n) \Big/ \binom{m+n}{m}.$$

Received June 29, 1954.

[1] This investigation was supported (in part) by research grant from The National Institutes of Health, Public Health Service.

[2] According to a review in *Mathematical Tables and Other Aids to Computation*, Vol. 6 (1952), p. 157, this table has been extended to $n = 10$ with $7D$ by H. R. van der Vaart. His table does not seem to be widely available in this country.

Simple recursion formulas permit the ready tabulation of A, but the problem of publication is formidable. The usefulness of a triple-entry table of exact values of A over the range of interest would scarcely justify the many pages it would require.[3]

Wilcoxon [1] presented without proof a formula which, for small values of u, permits one to obtain values of A from those of the double-entry quantity $A_0(u, m) = A(u, m, \infty)$. This function A_0 was studied and tabulated by Euler [5]. [More precisely, Euler tabled $a_0(u, m) = A_0(u, m) - A_0(u - 1, m)$, which is the number of ways of partitioning the exact value u into m parts.] In Section 2 we derive an identity similar in nature to that of Wilcoxon, but valid for all values of u. We also present tables of A_0 and of a related quantity A_2, from which values of A are readily obtained.

Our tables may be used provided $m \leqq 12$. This requirement that the smaller sample size not exceed 12 is considerably less restrictive than that the larger sample size not exceed 8, but still will leave many situations of interest uncovered. We turn therefore to approximations, and develop in Section 3 a polynomial expression for the sixth central moment of U. This permits us to obtain simple formulas for the coefficients of the Edgeworth series for π to terms of order $1/m^2$. A numerical investigation indicates this series to be reliable to about $4D$ when $m = 12$.

2. A combinatorial identity. To simplify notation, we adopt the conventions that $A(u, m, n)$ and $A_0(u, m)$ are 0 when $u < 0$, and that all variables of summation are integers. We observe

$$A(u, m, n) = A_0(u, m) - \sum_{t>n} A(u - t, \; m - 1, \; t). \tag{1}$$

This formula may be verified by observing that $A_0(u, m)$ counts the partitions $S_1 + S_2 + \cdots + S_m \leqq u$ where $0 \leqq S_1 \leqq \cdots \leqq S_m$, while $A(u, m, n)$ counts those of these partitions satisfying the additional restriction $S_m \leqq n$. Since $A(u - t, \; m - 1, \; t)$ is the number of the partitions with $S_m = t$, the sum in (1) represents just the number of partitions counted by $A_0(u, m)$ but not counted by $A(u, m, n)$.

We now apply (1) to itself repeatedly, obtaining the development

$$\begin{aligned} A(u, m, n) = A_0(u, m) &- \sum_{t_1>n} A_0(u - t_1, \; m - 1) \\ &+ \sum_{t_2>t_1>n} A_0(u - t_1 - t_2, \; m - 2) \\ &- \sum_{t_3>t_2>t_1>n} A_0(u - t_1 - t_2 - t_3, \; m - 3) + \cdots . \end{aligned} \tag{2}$$

This formula may now be simplified by the change of summation variable $s_i = t_i - n - i$. If we write $u - kn - \frac{1}{2}k(k + 1) = w$, the $(k + 1)$st term on

[3] Auble has attacked this problem by placing a table covering $m, n \leqq 20$ on file with the American Documentation Institute, from whom it may be purchased for \$4.25 (microfilm) or \$12.50 (photostat). See [4], p. 14 for details.

the right side of (2) may be written

$$(-1)^k \sum_{0 \leqq s_1 \leqq \cdots \leqq s_k} A_0[w - (s_1 + \cdots + s_k), \quad m - k].$$

In this sum, the term $A_0(w - v, \quad m - k)$ occurs as many times as there are ways of partitioning v into just k nonnegative integers, that is, $a_0(v, k)$ times. The $(k + 1)$st term of (2) is thus equal to

$$(-1)^k \sum_{v \geqq 0} a_0(v, k)\, A_0(w - v, \quad m - k).$$

If we now write

$$A_k(u, m) = \sum_{v \geqq 0} a_0(v, k)\, A_0(u - v, \quad m), \tag{3}$$

we can present (2) in the form

$$\begin{aligned} A(u, m, n) &= \sum_{k \geqq 0} (-1)^k A_k(u - kn - \tfrac{1}{2}k(k + 1), \quad m - k) \\ &= A_0(u, m) - A_1(u - n - 1, \quad m - 1) \\ &\quad + A_2(u - 2n - 3, \quad m - 2) - A_3(u - 3n - 6), \quad m - 3) + \cdots . \end{aligned} \tag{4}$$

The series is extended until the first argument becomes negative. Formulas (3) and (4) express the restricted partition function A in terms of the unrestricted partition function A_0.

We present in Table I the values of $A_0(u, m)$ for $m \leqq 12$ and $u \leqq 100$ [Euler's table of a_0 covers $m \leqq 20$ and $u \leqq 59$]. These values were computed with the aid of the familiar recursion relation

$$A_0(u, m) = A_0(u, \quad m - 1) + A_0(u - m, \quad m),$$

together with the boundary values $A_0(0, m) = 1$ and $A_0(u, 1) = u + 1$. Values of A_k for $k > 0$ can be computed from the relation

$$\begin{aligned} A_k(u, \quad m - k) &= \sum_{v=0}^{u} A_0(v, k)\, A_0(u - v, \quad m - k) \\ &\quad - \sum_{v=0}^{u} A_0(v - 1, \quad k)\, A_0(u - v, \quad m - k), \end{aligned}$$

but for convenience we also give in Table II the values of $A_2(u, m)$ for $m \leqq 11$ and $u \leqq 75$. Table II was computed with the aid of

$$A_2(u, m) = A_2(u, \quad m - 1) + A_2(u - m, \quad m), \qquad A_2(0, m) = 1,$$

which follow from (3). In the use of (4) for the range covered by our tables, one often needs A_1 and occasionally A_3. These quantities are readily obtained from Table II with the aid of

$$\begin{aligned} A_1(u, m) &= A_2(u, m) - A_2(u - 2, \quad m) \\ A_3(u, m) &= A_2(u, m) + A_2(u - 3, \quad m) + A_2(u - 6, \quad m) + \cdots . \end{aligned} \tag{5}$$

TABLE I

$A_0(u, m)$

u	$m = 2$	3	4	5	6	7	8	9	10	11	12
0	1	1	1	1	1	1	1	1	1	1	1
1	2	2	2	2	2	2	2	2	2	2	2
2	4	4	4	4	4	4	4	4	4	4	4
3	6	7	7	7	7	7	7	7	7	7	7
4	9	11	12	12	12	12	12	12	12	12	12
5	12	16	18	19	19	19	19	19	19	19	19
6	16	23	27	29	30	30	30	30	30	30	30
7	20	31	38	42	44	45	45	45	45	45	45
8	25	41	53	60	64	66	67	67	67	67	67
9	30	53	71	83	90	94	96	97	97	97	97
10	36	67	94	113	125	132	136	138	139	139	139
11	42	83	121	150	169	181	188	192	194	195	195
12	49	102	155	197	227	246	258	265	269	271	272
13	56	123	194	254	298	328	347	359	366	370	372
14	64	147	241	324	388	433	463	482	494	501	505
15	72	174	295	408	498	564	609	639	658	670	677
16	81	204	359	509	634	728	795	840	870	889	901
17	90	237	431	628	797	929	1025	1092	1137	1167	1186
18	100	274	515	769	996	1177	1313	1410	1477	1522	1552
19	110	314	609	933	1231	1477	1665	1803	1900	1967	2012
20	121	358	717	1125	1513	1841	2099	2291	2430	2527	2594
21	132	406	837	1346	1844	2277	2624	2889	3083	3222	3319
22	144	458	973	1601	2235	2799	3262	3621	3890	4085	4224
23	156	514	1123	1892	2689	3417	4026	4508	4874	5145	5340
24	169	575	1292	2225	3221	4150	4945	5584	6078	6448	6720

25	182	640	1477	2602	3833	5010	6035	6875	7533	8034	8406
26	196	710	1683	3029	4542	6019	7332	8424	9294	9964	10469
27	210	785	1908	3509	5353	7194	8859	10269	11406	12295	12972
28	225	865	2157	4049	6284	8561	10660	12463	13940	15107	16008
29	240	950	2427	4652	7341	10140	12764	15055	16955	18477	19663
30	256	1041	2724	5326	8547	11964	15226	18115	20545	22512	24064
31	272	1137	3045	6074	9907	14057	18083	21704	24787	27314	29326
32	289	1239	3396	6905	11447	16457	21402	25910	29800	33022	35616
33	306	1347	3774	7823	13176	19195	25230	30814	35688	39773	43092
34	324	1461	4185	8837	15121	22315	29647	36522	42600	47745	51969
35	342	1581	4626	9952	17293	25854	34713	43137	50670	57118	62458
36	361	1708	5104	11178	19725	29865	40525	50794	60088	68122	74842
37	380	1841	5615	12520	22427	34391	47155	59618	71024	80988	89394
38	400	1981	6166	13989	25436	39493	54719	69774	83714	96009	1 06478
39	420	2128	6754	15591	28767	45224	63307	81422	98377	1 13484	1 26456
40	441	2282	7386	17338	32459	51654	73056	94760	1 15305	1 33782	1 49790
41	462	2443	8058	19236	36529	58844	84074	1 09984	1 34771	1 57283	1 76946
42	484	2612	8778	21298	41023	66877	96524	1 27338	1 57138	1 84452	2 08516
43	506	2788	9542	23531	45958	75823	1 10536	1 47058	1 82746	2 15768	2 45094
44	529	2972	10358	25949	51385	85776	1 26301	1 69438	2 12038	2 51811	2 87427
45	552	3164	11222	28560	57327	96820	1 43975	1 94769	2 45439	2 93184	3 36276
46	576	3364	12142	31378	63837	1 09061	1 63780	2 23398	2 83486	3 40604	3 92573
47	600	3572	13114	34412	70941	1 22595	1 85902	2 55676	3 26700	3 94822	4 57280
48	625	3789	14147	37678	78701	1 37545	2 10601	2 92023	3 75737	4 56725	5 31567
49	650	4014	15236	41185	87143	1 54020	2 38094	3 32854	4 31231	5 27240	6 16634
50	676	4248	16390	44950	96335	1 72158	2 68682	3 78666	4 93971	6 07455	7 13933
51	702	4491	17605	48983	1 06310	1 92086	3 02622	4 29960	5 64731	6 98513	8 24969
52	729	4743	18890	53302	1 17139	2 13959	3 40260	4 87318	6 44456	8 01739	9 51529
53	756	5004	20240	57918	1 28859	2 37920	3 81895	5 51333	7 34079	9 18531	10 95477
54	784	5275	21665	62850	1 41551	2 64146	4 27926	6 22695	8 34733	10 50501	12 59017

TABLE—I *Continued*

$A_0(u, m)$

u	$m = 2$	3	4	5	6	7	8	9	10	11	12
55	812	5555	23160	68110	1 55253	2 92798	4 78700	7 02098	9 47537	11 99348	14 44442
56	841	5845	24735	73718	1 70053	3 24073	5 34674	7 90350	10 73836	13 67020	16 54447
57	870	6145	26385	79687	1 85997	3 58155	5 96249	8 88272	12 14972	15 55576	18 91852
58	900	6455	28120	86038	2 03177	3 95263	6 63945	9 96799	13 72536	17 67358	21 59931
59	930	6775	29935	92785	2 21644	4 35603	7 38225	11 16891	15 48122	20 04847	24 62127
60	961	7106	31841	99951	2 41502	4 79422	8 19682	12 49642	17 43613	22 70853	28 02420
61	992	7447	33832	1 07550	2 62803	5 26949	9 08844	13 96162	19 60893	25 68348	31 84982
62	1024	7799	35919	1 15606	2 85659	5 78457	10 06383	15 57716	22 02172	29 00685	36 14618
63	1056	8162	38097	1 24135	3 10132	6 34205	11 12905	17 35600	24 69679	32 71418	40 96387
64	1089	8536	40377	1 33162	3 36339	6 94494	12 29168	19 31266	27 65999	36 84530	46 36059
65	1122	8921	42753	1 42704	3 64348	7 59611	13 55860	21 46210	30 93747	41 44248	52 39725
66	1156	9318	45237	1 52787	3 94289	8 29892	14 93837	23 82109	34 55945	46 55293	59 14310
67	1190	9726	47823	1 63429	4 26232	9 05654	16 43879	26 40678	38 55650	52 22670	66 67112
68	1225	10146	50523	1 74658	4 60317	9 87266	18 06948	29 23839	42 96375	58 51951	75 06398
69	1260	10578	53331	1 86493	4 96625	10 75082	19 83926	32 33568	47 81690	65 49048	84 40900
70	1296	11022	56259	1 98963	5 35302	11 69507	21 75890	35 72052	53 15665	73 20512	94 80443
71	1332	11478	59301	2 12088	5 76436	12 70930	23 83835	39 41551	59 02444	81 73297	106 35424
72	1369	11947	62470	2 25899	6 20188	13 79799	26 08967	43 44567	65 46739	91 15087	119 17507
73	1406	12428	65759	2 40417	6 66649	14 96541	28 52401	47 83667	72 53346	101 54031	133 39013
74	1444	12922	69181	2 55674	7 15991	16 21645	31 15482	52 61692	80 27691	112 99109	149 13727
75	1482	13429	72730	2 71693	7 68318	17 55584	33 99463	57 81572	88 75319	125 59849	166 56236
76	1521	13949	76419	2 88507	8 23809	18 98891	37 05839	63 46517	98 02462	139 46710	185 82769
77	1560	14482	80241	3 06140	8 82576	20 52083	40 36009	69 59848	108 15498	154 70791	207 10516
78	1600	15029	84210	3 24627	9 44815	22 15745	43 91635	76 25203	119 21578	171 44248	230 58558
79	1640	15589	88319	3 43993	10 10642	23 90441	47 74276	83 46328	131 28018	189 79969	256 47081

80	1681	16163	92582	3 64275	10 80266	25 76807	51 85774	91 27325	144 42990	209 92038	284 98436
81	1722	16751	96992	3 85499	11 53817	27 75462	56 27863	99 72430	158 74874	231 95386	316 36286
82	1764	17353	1 01563	4 07703	12 31512	29 87096	61 02578	108 86245	174 32984	256 06281	350 86724
83	1806	17969	1 06288	4 30915	13 13491	32 12382	66 11845	118 73537	191 26883	282 41970	388 77394
84	1849	18600	1 11182	4 55175	13 99990	34 52073	71 57912	129 39484	209 67175	311 21206	430 38713
85	1892	19245	1 16237	4 80512	14 91154	37 06899	77 42908	140 89425	229 64744	342 63853	476 02866
86	1936	19905	1 21468	5 06967	15 87233	39 77674	83 69309	153 29157	251 31619	376 91468	526 05195
87	1980	20580	1 26868	5 34571	16 88388	42 65195	90 39471	166 64674	274 80172	414 26882	580 83118
88	2025	21270	1 32452	5 63367	17 94879	45 70341	97 56115	181 02443	300 24021	454 94812	640 77581
89	2070	21975	1 38212	5 93387	19 06878	48 93974	105 21837	196 49162	327 77180	499 21428	706 31944
90	2116	22696	1 44164	6 24676	20 24666	52 37048	113 39626	213 12056	357 55046	547 35015	777 93573
91	2162	23432	1 50300	6 57267	21 48421	56 00494	122 12339	230 98584	389 73458	599 65496	856 12577
92	2209	24184	1 56636	6 91207	22 78440	59 85339	131 43251	250 16788	424 49772	656 45158	941 43594
93	2256	24952	1 63164	7 26531	24 14919	63 92593	141 35501	270 74985	462 01868	718 08149	1034 44435
94	2304	25736	1 69900	7 63287	25 58166	68 23361	151 92670	292 82095	502 49270	784 91240	1135 77964
95	2352	26536	1 76836	8 01512	27 08390	72 78731	163 18202	316 47359	546 12103	857 33309	1246 10703
96	2401	27353	1 83989	8 41256	28 65922	77 59896	175 16011	341 80685	593 12304	935 76157	1366 14870
97	2450	28186	1 91350	8 82557	30 30978	82 68026	187 89863	368 92306	643 72478	1020 63946	1496 66812
98	2500	29036	1 98936	9 25467	32 03907	88 04401	201 44027	397 93189	698 17210	1112 44092	1638 49287
99	2550	29903	2 06739	9 70026	33 84945	93 70284	215 82623	428 94679	756 71859	1211 66671	1792 49789
100	2601	30787	2 14776	10 16288	35 74454	99 67047	231 10298	462 08882	819 63928	1318 85356	1959 62937

TABLE II

$A_2(u, m)$

u	$m = 1$	2	3	4	5	6	7	8	9	10	11
0	1	1	1	1	1	1	1	1	1	1	1
1	3	3	3	3	3	3	3	3	3	3	3
2	7	8	8	8	8	8	8	8	8	8	8
3	13	16	17	17	17	17	17	17	17	17	17
4	22	30	33	34	34	34	34	34	34	34	34
5	34	50	58	61	62	62	62	62	62	62	62
6	50	80	97	105	108	109	109	109	109	109	109
7	70	120	153	170	178	181	182	182	182	182	182
8	95	175	233	267	284	292	295	296	296	296	296
9	125	245	342	403	437	454	462	465	466	466	466
10	161	336	489	594	656	690	707	715	718	719	719
11	203	448	681	851	959	1021	1055	1072	1080	1083	1084
12	252	588	930	1197	1375	1484	1546	1580	1597	1605	1608
13	308	756	1245	1648	1932	2113	2222	2284	2318	2335	2343
14	372	960	1641	2235	2672	2964	3146	3255	3317	3351	3368
15	444	1200	2130	2981	3637	4091	4386	4568	4677	4739	4773
16	525	1485	2730	3927	4886	5576	6038	6334	6516	6625	6687
17	615	1815	3456	5104	6479	7500	8207	8672	8968	9150	9259
18	715	2200	4330	6565	8497	9981	11036	11751	12217	12513	12695
19	825	2640	5370	8351	11023	13136	14682	15754	16472	16938	17234
20	946	3146	6602	10529	14166	17130	19352	20932	22012	22731	23197
21	1078	3718	8048	13152	18038	22129	25275	27559	29156	30239	30958
22	1222	4368	9738	16303	22782	28358	32744	35999	38317	39922	41006
23	1378	5096	11698	20049	28546	36046	42084	46652	49969	52304	53912
24	1547	5915	13963	24492	35515	45496	53703	60037	64714	68065	70408

25	1729	6825	16563	29715	43881	57017	68053	76725	83241	87980	91348
26	1925	7840	19538	35841	53879	71009	85691	97442	1 06410	1 13035	1 17808
27	2135	8960	22923	42972	65754	87883	1 07235	1 22989	1 35206	1 44356	1 51043
28	2360	10200	26763	51255	79801	1 08159	1 33434	1 54366	1 70838	1 83351	1 92610
29	2600	11560	31098	60813	96328	1 32374	1 65118	1 92677	2 14689	2 31627	2 44322
30	2856	13056	35979	71820	1 15701	1 61197	2 03281	2 39280	2 68436	2 91167	3 08401
31	3128	14688	41451	84423	1 38302	1 95319	2 49022	2 95674	3 33991	3 64230	3 87427
32	3417	16473	47571	98826	1 64580	2 35589	3 03642	3 63679	4 13648	4 53570	4 84528
33	3723	18411	54390	1 15203	1 95004	2 82887	3 68578	4 45303	5 10017	5 62321	6 03327
34	4047	20520	61971	1 33791	2 30119	3 38278	4 45513	5 42955	6 26196	6 94261	7 48173
35	4389	22800	70371	1 54794	2 70495	4 02869	5 36303	6 59292	7 65702	8 53682	9 24090
36	4750	25270	79660	1 78486	3 16788	4 77985	6 43103	7 97469	9 32675	10 45710	11 37058
37	5130	27930	89901	2 05104	3 69684	5 65003	7 68284	9 60961	11 31799	12 76155	13 93963
38	5530	30800	1 01171	2 34962	4 29966	6 65555	9 14577	11 53857	13 68546	15 51897	17 02940
39	5950	33880	1 13540	2 68334	4 98453	7 81340	10 84982	13 80656	16 49092	18 80719	20 73329
40	6391	37191	1 27092	3 05578	5 76073	9 14351	12 82929	16 46608	19 80599	22 71766	25 16088
41	6853	40733	1 41904	3 47008	6 63796	10 66665	15 12178	19 57481	23 71129	27 35359	30 43760
42	7337	44528	1 58068	3 93030	7 62714	12 40699	17 77002	23 19957	28 29974	32 83544	36 70971
43	7843	48576	1 75668	4 44002	8 73968	14 38971	20 82074	27 41366	33 67562	39 29883	44 14411
44	8372	52900	1 94804	5 00382	9 98835	16 64390	24 32674	32 30143	39 95845	46 90106	52 93433
45	8924	57500	2 15568	5 62576	11 38649	19 19989	28 34566	37 95527	47 28202	55 81884	63 30057
46	9500	62400	2 38068	6 31098	12 94894	22 09245	32 94227	44 48084	55 79883	66 25593	75 49683
47	10100	67600	2 62404	7 06406	14 69120	25 35785	38 18714	51 99370	65 67916	78 44071	89 81129
48	10725	73125	2 88693	7 89075	16 63043	29 03742	44 15920	60 62528	77 11620	92 63517	106 57480
49	11375	78975	3 17043	8 79619	18 78454	33 17425	50 94427	70 51908	90 32507	109 13226	126 16166
50	12051	85176	3 47580	9 78678	21 17327	37 81717	58 63791	81 83748	105 54877	128 26643	148 99972
51	12753	91728	3 80421	10 86827	23 81721	43 01710	67 34384	94 75750	123 05724	150 41083	175 57171
52	13482	98658	4 15701	12 04776	26 73896	48 83141	77 17707	109 47850	143 15412	175 98956	206 42716
53	14238	1 05966	4 53546	13 33165	29 96208	55 31993	88 26220	126 21747	166 17592	205 47475	242 18446
54	15022	1 13680	4 94101	14 72779	33 51233	62 54975	100 73689	145 21773	192 49975	239 40081	283 54492

TABLE II—*Continued*

n	$m = 1$	2	3	4	5	6	7	8	9	10	11
55	15834	1 21800	5 37501	16 24328	37 41655	70 59080	114 75000	166 74370	222 54253	278 36137	331 29570
56	16675	1 30355	5 83901	17 88677	41 70398	79 52115	130 46542	191 09070	256 76986	323 02579	386 32636
57	17545	1 39345	6 33446	19 66611	46 40507	89 42217	148 06008	218 57916	295 69536	374 13607	449 63290
58	18445	1 48800	6 86301	21 59080	51 55288	100 38429	167 72813	249 56561	339 89068	432 52585	522 33714
59	19375	1 58720	7 42621	23 66949	57 18182	112 50175	189 67882	284 43632	389 98509	499 11735	605 69215
60	20336	1 69136	8 02582	25 91259	63 32914	125 87889	214 14109	323 61959	446 67683	574 94326	701 10492
61	21328	1 80048	8 66349	28 32960	70 03358	140 62438	241 36127	367 57874	510 73286	661 14369	810 14341
62	22352	1 91488	9 34109	30 93189	77 33696	156 85811	271 60811	416 82584	583 00176	758 99132	934 56303
63	23408	2 03456	10 06038	33 72987	85 28275	174 70492	305 17034	471 91404	664 41379	869 88854	1076 31570
64	24497	2 15985	10 82334	36 73593	93 91775	194 30204	342 36212	533 45282	755 99535	995 39616	1237 58062
65	25619	2 29075	11 63184	39 96144	103 29058	215 79233	383 52046	602 09962	858 86948	1137 23085	1420 77577
66	26775	2 42760	12 48798	43 41987	113 45345	239 33234	429 01116	678 57677	974 27213	1297 29792	1628 59362
67	27965	2 57040	13 39374	47 12361	124 46057	265 08495	479 22604	763 66236	1103 55304	1477 68911	1864 01547
68	29190	2 71950	14 35134	51 08727	136 37002	293 22813	534 58940	858 20899	1248 19408	1680 71993	2130 35283
69	30450	2 87490	15 36288	55 32432	149 24207	323 94699	595 55510	963 13384	1409 81067	1908 92802	2431 26516
70	31746	3 03696	16 43070	59 85057	163 14115	357 44319	662 61353	1079 43937	1590 17223	2165 11549	2770 80764
71	33078	3 20568	17 55702	64 68063	178 13408	393 92641	736 28853	1208 20257	1791 20433	2452 34802	3153 45294
72	34447	3 38143	18 74431	69 83158	194 29215	433 62449	817 14495	1350 59777	2015 01156	2774 00288	3584 14629
73	35853	3 56421	19 99491	75 31923	211 68925	476 77420	905 78536	1507 88498	2263 88033	3133 76887	4068 33190
74	37297	3 75440	21 31142	81 16199	230 40406	523 63219	1002 85823	1681 43500	2540 30448	3535 70064	4612 01634
75	38779	3 95200	22 69631	87 37694	250 51809	574 46508	1109 05448	1872 71684	2846 98897	3984 21982	5221 80044

These relations are so simple to use that tabulation of A_1 and A_2 is unnecessary. Values of A_k for $k > 3$ are seldom required. In general,

$$A_k(u, m) = \textstyle\sum_{r\geq 0} A_{k-1}(u - rk, \quad m).$$

We illustrate the tables by computing $\pi(95, 12, 22)$. Using (5) and the tables, we find

$$\begin{aligned}
A_0(95, 12) &= 124{,}610{,}703, \\
A_1(72, 11) &= 358{,}414{,}629 - 277{,}080{,}764 = 81{,}333{,}865, \\
A_2(48, 10) &= 9{,}263{,}517, \\
A_3(23, \ 9) &= 49{,}969 + 22{,}012 + 8{,}968 + 3{,}317 \\
&\qquad + 1{,}080 + 296 + 62 + 8 = 85{,}712.
\end{aligned}$$

Combining, $A(95, 12, 22) = 52{,}454{,}643$; this is the exact number of partitions of $95 + \frac{1}{2}12\cdot 13 = 173$ into just 12 distinct parts between 1 and 34 inclusive. Since $\binom{34}{12} = 548{,}354{,}040$, we find $\pi(95, 12, 22) = 0.095\,658 \cdots$.

3. Approximations. Our tables provide values of $\pi(u, m, n)$ only for $m \leqq 12$ and $u \leqq 100$. As is shown below, the normal approximation at these limits is subject to sizable percentage errors. In the search of better approximations for $m > 12$, we turn to the Edgeworth series, which to terms of order $1/m^2$ is (taking advantage of the symmetry of U)

$$\pi(u, m, n) \doteq \Phi(x) + e_{m,n}^{(3)}\varphi^{(3)}(x) + e_{m,n}^{(5)}\varphi^{(5)}(x) + e_{m,n}^{(7)}\varphi^{(7)}(x), \tag{6}$$

where x is the normalized value of u. Using $E(U) = \frac{1}{2}mn$ and $\mu_2 = mn(m + n + 1) / 12$, and the usual continuity correction, we take

$$x = (u + \tfrac{1}{2} - \tfrac{1}{2}mn) \,/\, \sqrt{mn\,(m + n + 1) / 12}. \tag{7}$$

The Edgeworth coefficients are given by

$$\begin{aligned}
e_{m,n}^{(3)} &= \frac{1}{4!}\left(\frac{\mu_4}{\mu_2^2} - 3\right), \qquad e_{m,n}^{(5)} = \frac{1}{6!}\left(\frac{\mu_6}{\mu_2^3} - 15\,\frac{\mu_4}{\mu_2^2} + 30\right), \\
& \qquad\qquad\qquad\qquad\qquad\qquad e_{m,n}^{(7)} = \frac{35}{8!}\left(\frac{\mu_4}{\mu_2^2} - 3\right)^2,
\end{aligned} \tag{8}$$

where μ_k is the kth central moment of U. Mann and Whitney give

$$\mu_4 = \frac{mn(m + n + 1)}{240}\,[5(m^2n + mn^2) - 2(m^2 + n^2) + 3mn - 2(m + n)].$$

They show [their formula (14)] that

$$\mu_6 = \frac{mn(m + n + 1)}{4032}\,[35m^2n^2(m^2 + n^2) + 70m^3n^3 + P(m, n)] \tag{9}$$

where $P(m, n)$ is a symmetric polynomial of 5th degree in m and n. When $m = 1$ and $m = 2$, the distribution may be given explicitly and the moments deter-

mined. In this way it may be shown that

$$
\begin{aligned}
P(m, n) = & -42\, mn(m^3 + n^3) - 14\, m^2n^2(m + n) + 16(m^4 + n^4) \\
(10) \qquad & -52\, mn(m^2 + n^2) - 43\, m^2n^2 + 32(m^3 + n^3) \\
& +14\, mn(m + n) + 8(m^2 + n^2) + 16\, mn - 8(m + n).
\end{aligned}
$$

If we substitute (10), (9) and (8) into (6) we find after simplification

$$
\begin{aligned}
\pi(u, m, n) \doteq\ & \Phi(x) - \frac{m^2 + n^2 + mn + m + n}{20mn(m + n + 1)} \varphi^{(3)}(x) \\
(11) \qquad & + \frac{\begin{array}{c}[2(m^4 + n^4) + 4mn(m^2 + n^2) + 6m^2n^2 + 4(m^3 + n^3) \\ + 7mn(m + n) + (m^2 + n^2) + 2mn - (m + n)]\end{array}}{210m^2n^2(m + n + 1)^2} \varphi^{(5)}(x) \\
& + \frac{(m^2 + n^2 + mn + m + n)^2}{800m^2n^2(m + n + 1)^2} \varphi^{(7)}(x).
\end{aligned}
$$

An appreciation of the accuracy of the Edgeworth approximations at the limit $m = 12$ of our tables may be gained from an examination of Table III. Column (a) gives the normal approximation; column (b) the first two terms of (11); column (c) the entire approximation (11); the last column gives the exact value.

TABLE III

m	n	u	(a)	(b)	(c)	$\pi(u, m, n)$
12	12	55	.17039	.17359	.17367	.17368
		45	.06301	.06380	.06388	.06384
		35	.01754	.01671	.01666	.01662
		25	.00363	.00285	.00280	.00278
12	24	100	.07218	.07303	.07308	.07307
		80	.01655	.01588	.01584	.01583
		60	.00254	.00202	.00199	.00198

It appears that (11) may be relied on to about $4D$ when $m = 12$, and its accuracy should improve with large values of m. The normal approximation (a) is subject to large percentage errors at the high significance levels, and is much improved by the use of the simple term in $\varphi^{(3)}(x)$.

REFERENCES

[1] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics*, Vol. 1 (1945), pp. 80–83.

[2] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 50–60.

[3] Colin White, "The use of ranks in a test of significance for comparing two treatments," *Biometrics*, Vol. 8 (1952), pp. 33–41.

[4] Donovan Auble, "Extended tables for the Mann-Whitney statistic," *Bulletin of the Institute of Educational Research, Indiana University*, Vol. 1, No. 2 (1953), 39 pp.

[5] L. Euler, "De partitione numerorum," *Opera Omnia, Series 1, Commentationes Arithmeticae*, Vol. 2, 1 (1915), pp. 254–294.

ON THE FOURIER SERIES EXPANSION OF RANDOM FUNCTIONS[1]

By W. L. Root and T. S. Pitcher

Massachusetts Institute of Technology

Problems involving stationary stochastic processes are often treated by approximating the original processes by Fourier series with orthogonal random coefficients.[2] In this paper we justify this technique in certain instances.

We let $x(t)$ denote a real- or complex-valued stochastic process defined for all values of t. We assume the first and second moments of $x(t)$ exist. We write $R(s, t) = E\{x(s)\overline{x(t)}\}$ and in case $R(s, t)$ is a function of $(t - s)$ only (that is, $x(t)$ is stationary in the wide sense), we write $\rho(t - s) = R(s, t)$. We assume everywhere that $E\{x(t)\} = 0$.

We define the stochastic process $x(t)$ to be *periodic* if the random variables $x(t_1)$ and $x(t_1 + T)$ are equal with probability one for all t_1 and some constant T. If $x(t)$ is periodic, then $R(s, t)$ is periodic in each variable. If $x(t)$ is wide-sense stationary, then it is periodic if and only if $\rho(\tau)$ is periodic.

Our first result follows from the theorem due independently to Karhunen and Loève which states: Let $x(t)$ be continuous in the finite interval (a, b), then

$$x(t) = \underset{n\to\infty}{\text{l.i.m.}} \sum_{1}^{n} x_i \psi_i(t)$$

where the $\psi_i(t)$ form an orthonormal system over (a, b) and where $E\{x_i\overline{x_j}\} = \lambda_i\delta_{ij}$ if and only if the $\psi_i(t)$ and the λ_i are a system of eigenfunctions and eigenvalues of the integral equation

$$\int_a^b R(s, t)\overline{\psi(t)}\, dt = \lambda\overline{\psi(s)}.$$

Theorem 1. *Let $x(t)$ be a wide-sense stationary stochastic process continuous in mean square. Then*

$$x(t) = \text{l.i.m.} \sum_{k=-n}^{n} \frac{x_i e^{ik\omega\tau}}{\sqrt{T}}, \qquad \omega = \frac{2\pi}{T},$$

on the interval $(0, T)$, where the x_i are pairwise orthogonal if and only if $x(t)$ is periodic with period T.

Proof. From the Karhunen-Loève theorem and the remark above about periodicity it follows that we need to show that

---

Received August 17, 1954.

[1] The research in this document was supported jointly by the Army, Navy, and Air Force under contract with the Massachusetts Institute of Technology.

[2] Of course, the use of Fourier series can usually be avoided by the use of the spectral representation theorem for stationary processes. See [1] p. 527.

$$\int_0^T R(s, t)\overline{\psi(t)}\, dt = \lambda\overline{\psi(s)}, \qquad 0 \leq s \leq T,$$

is satisfied by exponentials when and only when ρ is periodic with period T.

Suppose the ψ_j's are exponentials, that is

$$\psi_j(t) = e^{in_j\omega t} / \sqrt{T}, \qquad \omega = 2\pi/T,\ n_j \text{ an integer.}$$

Then by Mercer's theorem $\sum_j e^{in_j\omega t}e^{-in_j\omega s} / \lambda_j T$ converges uniformly on $[0, T] \times [0, T]$ to $R(s, t)$. Now since the process is stationary, $R(s, t) = \rho(s - t)$. Hence $\sum_j e^{in_j\omega\tau} / \lambda_j T$ converges uniformly to $\rho(\tau)$ in the interval $-T \leq \tau \leq T$. Thus $\rho(0) = \rho(T)$ and since $x(t)$ is wide-sense stationary, $E\{|x(t + T) - x(t)|^2\} = \rho(0) - 2R\,[\rho(T)] + \rho(0) = 0$. The converse is obvious.

Davis [2] shows that if $x(t)$ is a wide-sense stationary, continuous-in-mean-square process which has for every $T > 0$ a Fourier series, with random orthogonal coefficients, which converges to $x(t)$ in mean square over $(0, T)$, then $x(t)$ must be the trivial process with $\rho(\tau) = \text{const}$. This assertion follows from Theorem 1. For to satisfy Davis' hypothesis, ρ must satisfy $\rho(t) = \rho(t + T)$, for every $T > 0$, and the only functions with this property are constants.

A somewhat different problem from the expansion of a random function is the statistical representation of a random function. In particular, here we ask for conditions under which a sum of exponentials with orthogonal random coefficients has the same multivariate probability distributions as a given random function. If $x(t)$ is Gaussian, such a statistical representation essentially always exists.

THEOREM 2. *Let $x(t)$ be a stationary Gaussian process with correlation function*

$$\rho(\tau) = \underset{n\to\infty}{\text{l.i.m.}} \sum_{-n}^{n} c_k e^{i\omega k t}, \quad -\frac{T}{2} \leq t \leq \frac{T}{2}, \quad \omega = \frac{2\pi}{T}.$$

Then there exists a process $y(t)$ defined by

$$y(t) = \underset{n\to\infty}{\text{l.i.m.}} \sum_{-n}^{n} x_k e^{i\omega k t}$$

where the x_k are complex Gaussian variables satisfying

$$E\{x_k\} = 0, \quad E\{x_k\overline{x_k}\} = c_k, \quad E\{x_k\overline{x_j}\} = 0, \qquad k \neq j,\ x_k = \overline{x_{-k}},$$

which has the same multivariate distributions as $x(t)$ over the interval $0 \leq t \leq \frac{1}{2}T$.

PROOF. It is easily verified that if $y_n(t) = \sum_{-n}^{n} x_k e^{i\omega k t}$, then

$$\lim_{\substack{m\to\infty\\ n\to\infty}} E\{|y_m(t) - y_n(t)|^2\} = 0,$$

so the process $y(t)$ is defined. Then

$$E\{y(t)\overline{y(t+\tau)}\} = \sum_{k,n=-\infty}^{\infty} E\{x_k\overline{x_n}\}e^{iwkt}e^{-iwn(t+\tau)} = \sum_{-\infty}^{\infty}\{|x_k|^2\}e^{-ikw\tau} = \rho(-\tau),$$

which proves the assertion.

For any $T < \infty$, $\int_{-T}^{T} E(|x(t)|^2)\,dt = 2TE(|x(o)|^2) < \infty$, so that, by Fubini's theorem, $x(t)$ is almost always measurable and square integrable on $(-T, T)$. Hence for almost every sample function $x(t) = \text{l.i.m.} \sum c_n e^{in\pi t/T}$ on the interval $(-T, T)$, where $c_n = (2T)^{-1}\int_{-T}^{T} x(s)e^{\pi nis/T}\,ds$. The process defined by the two preceding equations has the same (a.e.) sample functions as the original on $(-T, T)$. We have

$$\begin{aligned} E(c_n\overline{c_m}) &= (2T)^{-2}\int_{-T}^{T} ds \int_{-T}^{T} dt\, e^{\pi i(ns-mt)/T}\rho(s-t) \\ &= (4T)^{-1}\int_{-1}^{1} dv\, e^{\pi i(n-m)v}\int_{(v-1)T}^{(v+1)T} du\, e^{im\pi u/T}\rho(u) \end{aligned} \tag{1}$$

where we have made the substitutions $u = s - t$ and then $v = s/T$. If $n = m$, then the integrand is dominated by $2T\rho(o)$, so that $E(|c_n|^2) \leqq \rho(o)$. That is, c_n is square integrable and $E(c_n\overline{c_m})$ exists.

THEOREM 3. *If ρ is integrable on $(-\infty, \infty)$ and $\int_{-\infty}^{\infty}\rho(t)\,dt \neq 0$, then $E(|c_n|^2) = O(1/T) > 0$. If also $n \neq m$, then*

$$\lim_{T\to\infty}\frac{E(c_n\overline{c_m})}{[E(|c_n|^2)E(|c_m|^2)]^{1/2}} = 0.$$

PROOF. The functions f_T defined by

$$f_T(v) = \int_{(v-1)T}^{(v+1)T} du\, e^{m\pi u/T}\rho(u)$$

are uniformly bounded by $\int_{-\infty}^{\infty}|\rho(u)|\,du$. For every v, with $|v| < 1$

$$f_T(v) \to \int_{-\infty}^{\infty}\rho(u)\,du = a \quad \text{as} \quad T\to\infty.$$

Hence, by Lebesgue's theorem applied to (1)

$$\lim\,(4T)E(c_n\overline{c_m}) = \begin{cases} 0 & n \neq m, \\ 2a & n = m. \end{cases}$$

This implies the result.

THEOREM 4. *If ρ is square integrable and r is its Fourier transform, then*

$$E(c_n\overline{c_m}) = \frac{(-1)^{n+m}}{4\pi T}\int_{-\infty}^{\infty} r\left(\frac{u+n\pi}{2\pi T}\right)\frac{\sin^2 u}{u[u+(n-m)\pi]}\,du. \tag{2}$$

PROOF.

$$E(c_n\overline{c_m}) = \frac{1}{4T}\int_{-1}^{1} dv\, e^{i\pi(n-m)v}\int_{(v-1)T}^{(v+1)T} du\, e^{i\pi mu/T}\left\{\underset{A\to\infty}{\text{l.i.m.}}\int_{-A}^{A} dw\, e^{-i2\pi uw}r(w)\right\}$$

$$= \frac{1}{4T}\int_{-1}^{1} dv\, e^{i\pi(n-m)v}\lim_{A\to\infty}\left\{\int_{(v-1)T}^{(v+1)T} du\, e^{i\pi mu/T}\int_{-A}^{A} dw\, e^{-i2\pi uw}r(w)\right\}$$

$$= \frac{(-1)^{m+1}}{4T}\int_{-1}^{1} dv\, e^{i\pi(n-m)v}\lim_{A\to\infty}\int_{-A}^{A} 2r(w)e^{i(m\pi/T-2\pi w)vT}\frac{\sin 2\pi wT\, dw}{(m\pi/T-2\pi w)}$$

$$= \lim_{A\to\infty}\frac{(-1)^{m+1}}{2T}\int_{-A}^{A} r(w)\frac{\sin 2\pi wT}{(m\pi/T-2\pi w)}\, dw\int_{-1}^{1} e^{i\pi v(n-2wT)}\, dv.$$

This yields the formula of the theorem. The first step is justified by an application of the Schwarz inequality, the second by the Fubini theorem, and the third by the Lebesgue bounded convergence theorem and the Fubini theorem.

THEOREM 5. *If ρ is square integrable on $(-\infty, \infty)$ and its Fourier transform vanishes almost everywhere in $(-\epsilon, \epsilon)$ for some $\epsilon > 0$, then*

$$\lim_{T\to\infty}\frac{E(c_n\overline{c_m})}{[E(|c_n|^2)E(|c_m|^2)]^{1/2}} = (-1)^{n-m}$$

for all n and m.

PROOF. By the hypothesis, the integrand in the previous theorem vanishes if $-2\pi\epsilon T - n\pi \leqq u \leqq 2\pi\epsilon T - n\pi$ and, for any $\theta < 1$, T can be chosen so large that outside this interval

$$\frac{\sin^2 u}{u[u+(n-m)\pi]} \geqq \frac{\theta\sin^2 u}{u^2}.$$

Since $r(x) \geqq 0$ for almost all x, for large T

$$r\left(\frac{u+n\pi}{2\pi T}\right)\frac{\sin^2 u}{u[u+(m-n)\pi]} \geqq \theta r\left(\frac{u+n\pi}{2\pi T}\right)\frac{\sin^2 u}{u^2}.$$

Integrating this gives $(-1)^{n+m}E(c_n\overline{c_m}) \geqq \theta E(|c_n|^2)$. Hence for large enough T,

$$1 \geqq \frac{(-1)^{n+m}E(c_n\overline{c_m})}{[E(|c_n|^2)E(|c_m|^2)]^{1/2}} \geqq \theta,$$

which implies the result.

Instead of holding n and m constant, we can fix the frequencies associated with them, that is, set $n = aT$ and $m = bT$.

THEOREM 6. *If ρ is integrable and square integrable, if a is a real number and $p \neq q$ are integers such that $r(a/2) > 0$ and $r((q/p)(a/2)) > 0$ then*

$$\lim_{T_k\to\infty}\frac{E(c_{pk}\overline{c_{qk}})}{[E(|c_{pk}|^2)E(|c_{qk}|^2)]^{1/2}} = 0, \qquad T_k = \frac{kp}{a}, \qquad k = 0, 1, 2, \cdots.$$

In fact the expression above is $O([(p-q)k]^{-\theta})$ for any $\theta < 1$.

PROOF. By Lebesgue's theorem and the boundedness of r, which follows from the integrability of ρ,

$$\frac{4\pi pk}{a} E(|c_{pk}|^2) = \int_{-\infty}^{\infty} r\left(\frac{au}{pk} + \frac{a}{2}\right) \frac{\sin^2 u}{u^2}\, du \to r\left(\frac{a}{2}\right) \int_{-\infty}^{\infty} \frac{\sin^2 u}{u^2}\, du = \pi r\left(\frac{a}{2}\right).$$

Similarly, $(4\pi p^2 k/qa)E(|c_{qk}|^2) \to r(\,(q/p)(a/2)\,\pi)$. Thus

$$\left|\frac{4\pi pk}{a} E(c_{pk}\,\overline{c_{qk}})\right| = \left|\int_{-\infty}^{\infty} r\left(\frac{au}{2\pi pk} + \frac{a}{2}\right) \frac{\sin^2 u}{u[u + (q-p)k\pi]}\, du\right|$$

$$\leqq \max |r(x)| \int_{-\infty}^{\infty} \left|\frac{\sin^2 u}{u[u + (q-p)k\pi]}\right| du.$$

But the above integral is $O(([q - p]k)^{-\theta})$ for any $\theta < 1$.

THEOREM 7. *If ρ is integrable and square integrable, a is a point at which $r(a/2) > 0$, and p and q are integers, then*

$$\lim_{T_k \to \infty} \frac{E(c_{pk}\,\overline{c_{pk+q}})}{[E(|c_{pk}|^2)E(|c_{pk+q}|^2)]^{1/2}} = 0, \qquad T_k = \frac{kp}{a}, \qquad k = 0, 1, 2, \cdots.$$

PROOF. As above, we get $\lim\, (4\pi pk/a)(E|c_{pk}|^2) = \lim\, (4\pi pk/a)E(|c_{pk+q}|^2) = r(a/2)\pi$. Now

$$\left|\frac{4\pi pk}{a} E(c_{pk}\,\overline{c_{pk+q}})\right| = \left|\int_{-\infty}^{\infty} r\left(\frac{au}{2\pi pk} + \frac{a}{2}\right) \frac{\sin^2 u}{u(u + q\pi)}\, du\right|$$

$$= \left|\int_{-\infty}^{\infty} \left[r\left(\frac{au}{2\pi pk} + \frac{a}{2}\right) - r\left(\frac{a}{2}\right)\right] \frac{\sin^2 u}{u(u + q\pi)}\, du\right|,$$

since $\int_{-\infty}^{\infty} (\sin^2 u)\, [u(u + q\pi)]^{-1}\, du = 0$ for $q \neq 0$. But by Schwarz's inequality

$$\left|\frac{4\pi pk}{a} E(c_{pk}\overline{c_{pk+q}})\right|^2 \leqq \int_{-\infty}^{\infty} \frac{\sin^2 v}{v(v + q\pi)^2}\, dv \int_{-\infty}^{\infty} \left|r\left(\frac{au}{2\pi pk} + \frac{a}{2}\right) - r\left(\frac{a}{2}\right)\right|^2 \frac{\sin^2 u}{u^2}\, du.$$

The second integral approaches zero by Lebesgue's theorem.

It is easy to find examples of processes, vanishing around $a/2$ and $b/2$, for which the conclusion of Theorem 6 is false. Suppose, for example, r vanishes in the intervals

$$\left[\left(\frac{a}{2}\right)\left(\frac{p - \epsilon q}{p - \epsilon p}\right),\ \left(\frac{a}{2}\right)\left(\frac{q}{p}\right)\right],$$

$$\left[\left(\frac{qa}{2p}\right)\left(\frac{p}{q}\right)^2,\ \left(\frac{qa}{2p}\right)\left(\frac{q - p\epsilon}{p - p\epsilon}\right)\right], \qquad q > p > 0, \epsilon > 0.$$

Then $\epsilon E(|c_{pk}|^2) - |E(c_{pk}\overline{c_{qk}})| < 0$ and $\epsilon E(|c_{qk}|^2) - |E(c_{qk}\overline{c_{pk}})| < 0$.

Since the coefficients corresponding to frequencies x with $r(x) = 0$ tend to misbehave, it is desirable to show that the total effect of such coefficients is small for large T. The following theorem does this for the band limited case.

THEOREM 8. *If r is bounded and vanishes outside $(-A/2, A/2)$ for some A, then*

$$\left[\sum_{[AT]+1}^{\infty} E(|c_n|^2) + \sum_{-[AT]+1}^{-\infty} E(|c_n|^2)\right] \Big/ \sum_{-\infty}^{\infty} E(|c_n|^2) = 0\left(\frac{\log T}{T^2}\right),$$

where $[AT]$ is the largest integer not exceeding AT.

PROOF.

$$\sum_{-\infty}^{\infty} E(|c_n|^2) = E\left(\int_{-T}^{T} |x(t)|^2\, dt\right) = \int_{-T}^{T} E(|x(t)|^2)\, dt = 2T\rho(o),$$

$$\sum_{[AT]+1}^{\infty} E(|c_n|^2) = \sum_{[AT]+1}^{\infty} \frac{1}{4\pi T} \int_{-\infty}^{\infty} r\left(\frac{u+n\pi}{2\pi T}\right) \frac{\sin^2 u}{u^2}\, du$$

$$= \frac{1}{2} \sum_{[AT]+1}^{\infty} \int_{-\infty}^{\infty} r(v) \frac{\sin^2 2\pi v T}{(2\pi v T + n\pi)^2}\, dv.$$

Now,

$$\sum_{[AT]+1}^{\infty} \frac{1}{(2\pi v T + n\pi)^2} = \frac{1}{\pi^2} \sum_{0}^{\infty} \frac{1}{(2vT + [AT] + 1 + n)^2}$$

$$= \frac{1}{\pi^2} (\log \Gamma)''(2vT + [AT] + 1),$$

$$\sum_{[AT]+1}^{\infty} E(|c_n|^2) \frac{1}{2\pi^2} \int_{-A/2}^{A/2} r(v) \sin^2 2\pi T v (\log \Gamma)''(2vT + [AT] + 1)\, dv$$

$$\leqq \frac{1}{2\pi^2} \sup (r) \int_{-A/2}^{A/2} (\log \Gamma)''(2vT + [AT] + 1)\, dv$$

$$= \frac{\sup (r)}{4\pi^2 T} \int_{[AT]+1-AT}^{[AT]+1+AT} (\log \Gamma)''(\omega)\, d\omega$$

$$\leqq \frac{\sup (r)}{4\pi^2 T} (\log \Gamma)'(\omega) \Big|_{[AT]-AT+1}^{[AT]+AT+1} = O\left(\frac{\log T}{T}\right).$$

The other case is handled similarly.

REFERENCES

[1] J. L. DOOB, *Stochastic Processes*, John Wiley and Sons, New York, 1953.

[2] R. C. DAVIS, "On the Fourier expansion of stationary random processes," *Proc. Amer. Math. Soc.*, Vol. 4 (1953), pp. 564–569.

A CHARACTERIZATION OF THE GAMMA DISTRIBUTION

By Eugene Lukacs

Office of Naval Research

1. Introduction. The sum and the difference of two independently and identically distributed normal variates are uncorrelated and therefore also independent. Conversely, we can conclude from the independence of the sum and the difference of two identically and independently distributed random variables that both these variables are normally distributed. In this manner one obtains a characterization of the normal distribution.

We denote in the following by

$$F(x; \alpha, \lambda) = \begin{cases} 0 & x < 0, \\ \dfrac{\alpha^\lambda}{\Gamma(\lambda)} \displaystyle\int_0^x t^{\lambda-1} e^{-\alpha t}\, dt & x > 0, \end{cases} \tag{1.1}$$

the distribution function of the gamma distribution. The corresponding characteristic function is

$$f(t; \alpha, \lambda) = (1 - it/\alpha)^{-\lambda}. \tag{1.2}$$

Here α and $\lambda > 0$ are two parameters. It is seen easily that α is a scale parameter.

Let X and Y be two identically and independently distributed random variables each having the distribution (1.1). It is known [2] that in this case $U = X + Y$ and $V = X/Y$ are also two independent random variables.

In the present paper, we use this fact to derive a characterization of the gamma distribution which is similar to the characterization of the normal distribution mentioned above. Our result can be formulated in the following manner.

Theorem. *Let X and Y be two nondegenerate and positive random variables, and suppose that they are independently distributed. The random variables $U = X + Y$ and $V = X/Y$ are independently distributed if and only if both X and Y have gamma distributions with the same scale parameter.*

In the following it will be convenient to introduce the random variable

$$W = 1/(1 + V) = Y/(X + Y). \tag{1.3}$$

The random variables U and W are both nonnegative. Moreover, W is a bounded random variable and

$$0 \leqq W \leqq 1. \tag{1.4}$$

2. Analytic properties of the characteristic functions of X, Y, U, and W. In this section we consider the characteristic functions of these random variables and investigate in particular whether the integrals defining these functions exist

Received July 2, 1954.

also for some complex values of the argument. In addition, we discuss some properties of these functions which we need in the subsequent sections.

We denote in the following by $F(x)$, $G(y)$, $H_1(u)$, and $H_2(w)$ the distribution functions of the random variables X, Y, U, and W respectively. Since X and Y are nonnegative, the characteristic functions of these random variables are

$$f(v) = \int_0^\infty e^{ivx}\, dF(x), \qquad g(v) = \int_0^\infty e^{ivy}\, dG(y). \tag{2.1}$$

Due to the nonnegativity of X and Y, the range of integration extends only from 0 to ∞. Therefore these integrals exist not only for real v but also for complex values $v = s + it$, where s and t are real, for which $t = \operatorname{Im}(v) \geqq 0$. It follows then from well-known properties of the Laplace integral that the functions (2.1) are analytic for $t = \operatorname{Im}(v) > 0$ and that

$$\begin{cases} f'(v) = i \int_0^\infty x e^{ivx}\, dF(x), & f''(v) = -\int_0^\infty x^2 e^{ivx}\, dF(x), \\ g'(v) = i \int_0^\infty y e^{ivy}\, dG(y), & g''(v) = -\int_0^\infty y^2 e^{ivy}\, dG(y). \end{cases} \tag{2.2}$$

We need also the following continuity property of the integrals (2.1),

$$\lim_{t \downarrow 0} f(s + it) = f(s); \qquad \lim_{t \downarrow 0} g(s + it) = g(s) \tag{2.3}$$

where $t \downarrow 0$ means that t approaches zero from above.

We consider next $\mathbf{E}\{e^{izW}\}$, the characteristic function of the random variable W. It is seen from (1.4) that

$$\mathbf{E}\{e^{izW}\} = \int_{-\infty}^{+\infty} e^{izw}\, dH_2(w) = \int_0^1 e^{izw}\, dH_2(w).$$

This integral exists for all values of z and is an entire function of z. The random variable U is nonnegative. Its characteristic function is

$$\mathbf{E}\{e^{ivU}\} = \int_{-\infty}^{\infty} e^{ivu}\, dH_1(u) = \int_0^\infty e^{ivu}\, dH_1(u).$$

Therefore this integral exists for all values of v for which $t = \operatorname{Im}(v) \geqq 0$ and is analytic for all v for which $t > 0$. Similarily the expectation $\mathbf{E}\{e^{ivU+izW}\}$ exists for all z and all v such that $t = \operatorname{Im}(v) \geqq 0$ and is analytic for all z and all v with $t = \operatorname{Im}(v) > 0$.

We conclude this section with another remark concerning W. The random variable W is bounded; therefore all its moments exist. We denote in the following by

$$\theta_1 = \mathbf{E}\left(\frac{Y}{X+Y}\right), \qquad \theta_2 = \mathbf{E}\left[\left(\frac{Y}{X+Y}\right)^2\right]. \tag{2.4}$$

It is then easy to see that

$$0 < \theta_1^2 \leqq \theta_2 \leqq \theta_1 < 1. \tag{2.5}$$

3. The independence of U and V. The random variables U and V are assumed to be independent. It follows then from (1.3) that this is equivalent to the independence of U and W. We denote by $H(u, w)$ the joint distribution of the random variables U and W and have then

$$H(u, w) = H_1(u)\, H_2(w). \tag{3.1}$$

We conclude from (3.1) that $\mathbf{E}[\exp(ivU + izW)] = \mathbf{E}[\exp(ivU)]\, \mathbf{E}[\exp(izW)]$, or

$$\begin{aligned}\mathbf{E}\left\{\exp\left[iv(X+Y) + \frac{izY}{X+Y}\right]\right\} \\ = \mathbf{E}\,\{\exp[iv(X+Y)]\}\, \mathbf{E}\left\{\exp\left[\frac{izY}{X+Y}\right]\right\}.\end{aligned} \tag{3.2}$$

We have shown in the preceding section that the expectations in (3.2) are analytic for all z and for all v such that $t = \mathrm{Im}(v) > 0$. We now write (3.2) in a more explicit form and obtain

$$\begin{aligned}\int_0^\infty\int_0^\infty \exp\left[iv(x+y) + \frac{izy}{x+y}\right] dF(x)\, dG(y) \\ = \int_0^\infty\int_0^\infty \exp[iv(x+y)]\, dF(x)\, dG(y) \int_0^\infty\int_0^\infty \exp\left[\frac{izy}{x+y}\right] dF(x)\, dG(y).\end{aligned} \tag{3.3}$$

The integrals in (3.3) converge for all complex z and all $v = s + it$ such that $t \geqq 0$. Therefore they are analytic in z and also in v, provided that $t = \mathrm{Im}(v) > 0$.

Equation (3.3) is the starting point for our investigation. We shall derive from it two relations for the unknown characteristic functions.

4. The relation between the characteristic functions. We assume for the time being that $t > 0$. Therefore we can differentiate (3.3), first with respect to v and then with respect to z. In this manner we obtain

$$\begin{aligned}\int_0^\infty\int_0^\infty y \exp\left[iv(x+y) + \frac{izy}{x+y}\right] dF(x)\, dG(y) \\ = \int_0^\infty\int_0^\infty (x+y) \exp[iv(x+y)]\, dF(x)\, dG(y) \\ \cdot \int_0^\infty\int_0^\infty \frac{y}{x+y} \exp\left[\frac{izy}{x+y}\right] dF(x)\, dG(y).\end{aligned} \tag{4.1}$$

If we put here $z = 0$ and use the notation (2.4), we see that, if $t = \mathrm{Im}(v) > 0$,

$$\begin{aligned}\int_0^\infty\int_0^\infty y \exp[iv(x+y)]\, dF(x)\, dG(y) \\ = \theta_1 \int_0^\infty\int_0^\infty (x+y) \exp[iv(x+y)]\, dF(x)\, dG(y).\end{aligned} \tag{4.2}$$

We use (2.1) to express the integrals in (4.2) in terms of the functions $f(v)$ and $g(v)$ and obtain

$$(4.3) \qquad (1 - \theta_1)\, g'(v)\, f(v) = \theta_1 f'(v)\, g(v), \qquad t = \operatorname{Im}(v) > 0.$$

From (2.3) and from

$$(4.4) \qquad f(0) = g(0) = 1$$

we conclude that there exists a neighborhood of the origin such that

$$(4.5) \qquad f(v) \neq 0, \qquad g(v) \neq 0,$$

for all values of v belonging to this neighborhood for which $t = \operatorname{Im}(v) \geqq 0$. (This neighborhood could, of course, be the half-plane $t \geqq 0$.) Moreover, it follows from (2.1) that $f(it)$ and $g(it)$ are positive for real $t > 0$. From the continuity of $f(v)$ and $g(v)$ we see that every point of the segment $0 < \operatorname{Im}(v) \leqq 1$ has a neighborhood in which (4.5) holds. We conclude then from the Heine-Borel covering theorem that there exists a simply connected domain $\mathfrak{D}$ containing the interval $0 < \operatorname{Im}(v) \leqq 1$ in which $f(v)$ and $g(v)$ do not vanish. In the following we restrict ourselves to this domain. We may then divide (4.3) by $f(v)g(v)$ and obtain

$$(4.6) \qquad (1 - \theta_1)(g'(v)/g(v)) = \theta_1(f'(v)/f(v)).$$

In the domain $\mathfrak{D}$ we may introduce the logarithms of $f(v)$ and $g(v)$ and integrate (4.6) using the initial conditions (4.4). We obtain finally

$$(4.7) \qquad [g(v)]^{1-\theta_1} = [f(v)]^{\theta_1}.$$

This relation is certainly valid for values of v for which the relations (4.5) hold together with

$$(4.8) \qquad t = \operatorname{Im} v > 0.$$

5. The differential equation. We derived relation (4.7) which connects the two unknown functions $f(v)$ and $g(v)$. To determine these function we must find a second relation. This may be accomplished by repeating the procedure which led from (4.1) to (4.6). We still restrict our considerations to values of v from $\mathfrak{D}$ for which (4.5) and (4.8) hold. We may therefore differentiate (4.1) twice, first with respect to v then with respect to z. We put finally $z = 0$ and see, using again the notations of (2.4), that

$$\int_0^\infty\int_0^\infty y^2 \exp\,[i(x + y)v]\, dF(x)\, dG(y)$$
$$= \theta_2 \int_0^\infty\int_0^\infty (x^2 + 2xy + y^2) \exp\,[iv(x + y)]\, dF(x)\, dG(y).$$

If we substitute for these integrals the expressions (2.1) and (2.2), we obtain a second relation between the functions $f(v)$ and $g(v)$,

$$g''(v)\, f(v) = \theta_2[f''(v)\, g(v) + 2f'(v)\, g'(v) + g''(v)\, f(v)].$$

On account of (4.5) we may write this in the form

$$\frac{g''(v)}{g(v)} = \theta_2 \left[\frac{f''(v)}{f(v)} + 2\frac{f'(v)}{f(v)}\frac{g'(v)}{g(v)} + \frac{g''(v)}{g(v)} \right]. \tag{5.1}$$

It is now convenient to introduce the logarithms

$$\phi(v) = \log f(v), \qquad \psi(v) = \log g(v). \tag{5.2}$$

Then

$$\begin{cases} \dfrac{f'(v)}{f(v)} = \phi'(v), & \dfrac{g'(v)}{g(v)} = \psi'(v) \\[2ex] \dfrac{f''(v)}{f(v)} = \phi''(v) + [\phi'(v)]^2, & \dfrac{g''(v)}{g(v)} = \psi''(v) + [\psi'(v)]^2. \end{cases} \tag{5.3}$$

From (4.6) and (5.2) we see that

$$(1 - \theta_1)\psi'(v) = \theta_1 \phi'(v), \qquad (1 - \theta_1)\psi''(v) = \theta_1 \phi''(v). \tag{5.4}$$

After some elementary computations we obtain from (5.1), (5.3), and (5.4) the differential equation

$$(1 - \theta_1)(\theta_1 - \theta_2)\phi''(v) = (\theta_2 - \theta_1^2)[\phi'(v)]^2. \tag{5.5}$$

6. Solution of the differential equation. We first leave aside the cases where $\theta_1 = \theta_2$ or $\theta_2 = \theta_1^2$ and consider the case where $0 < \theta_1^2 < \theta_2 < \theta_1 < 1$. Then

$$\rho = \frac{(1 - \theta_1)(\theta_1 - \theta_2)}{\theta_2 - \theta_1^2} > 0 \tag{6.1}$$

and we may write (5.5) as

$$\phi''(v)/[\phi'(v)]^2 = 1/\rho \tag{6.2}$$

for values of v satisfying (4.5) and (4.8). We denote by

$$k_1 = \mathbf{E}(e^{-X}), \qquad k_2 = \mathbf{E}(Xe^{-X}), \qquad \alpha = (k_1\rho - k_2)/k_2 \tag{6.3}$$

and integrate (6.2) with the initial condition $\phi'(i) = ik_2/k_1$. We obtain easily

$$\frac{1}{\phi'(v)} = -\frac{v}{\rho} + \frac{k_1\rho - k_2}{i\rho k_2} = -\frac{v}{\rho} + \frac{\alpha}{i\rho}$$

so that $\phi'(v) = (i\rho/\alpha)/(1 - iv/\alpha)$. We integrate this again and obtain, considering (5.2),

$$f(v) = c_2(1 - iv/\alpha)^{-\rho} \tag{6.4}$$

where c_2 is a constant of integration. It follows from (2.3) that $\lim_{t\downarrow 0} f(it) = f(0) = 1$, therefore $c_2 = 1$. It is then seen from (6.4) and (4.7) that

$$f(v) = (1 - iv/\alpha)^{-\rho}, \qquad g(v) = (1 - iv/\alpha)^{-\theta_1\rho/(1-\theta_1)}. \tag{6.5}$$

The equations (6.5) were derived for all points v of the domain $\mathfrak{D}$. This restriction on v can now be removed. The functions $f(v) = \int_0^\infty e^{itv}\, dF(x)$ and $(1 - iv/\alpha)^{-\rho}$

are both analytic in the upper half-plane $t = \text{Im}(v) > 0$, and agree in the domain $\mathfrak{D}$. Hence they agree in the half-plane $\text{Im}(v) > 0$. The same argument applies to $g(v)$. Finally it follows from (2.3) that (6.5) holds also for real values of v. The characteristic functions of X and Y are therefore given by (6.5) so that X and Y have, indeed, gamma distributions with the same scale parameter. We still have to consider the cases $\theta_1 = \theta_2$ and $\theta_1^2 = \theta_2$. If $\theta_1 = \theta_2$ then we see from (2.5) that $\theta_2 > \theta_1^2$, therefore (5.5) reduces to $\phi'(v) = 0$. If on the other hand $\theta_1^2 = \theta_2$, then we obtain from (5.5) the equation $\phi''(v) = 0$. By a reasoning similar to the one we employed earlier, it is seen that in both these cases the random variables X and Y have degenerate distributions.

Therefore, we have established that the independence of U and V implies that X and Y have gamma distributions with the same scale parameter. We still have to prove the converse. This does not follow from Pitman's result [2] quoted in the introduction, since we did not assume that X and Y have the same distribution.

Suppose, therefore, that X and Y are independent random variables and that their distributions are $F(x; \alpha, \lambda_1)$ and $F(x; \alpha, \lambda_2)$, respectively. We see, then, from (1.1) that the characteristic function of the joint distribution of U and V is given by

$$\mathbf{E}\left\{\exp\left[it(X+Y) + \frac{izX}{Y}\right]\right\}$$
$$= \frac{\alpha^{\lambda_1+\lambda_2}}{\Gamma(\lambda_1)\Gamma(\lambda_2)} \int_0^\infty\int_0^\infty \exp\left[-(\alpha - it)(x+y) + \frac{izx}{y}\right] x^{\lambda_1-1} y^{\lambda_2-1}\,dx\,dy.$$

If we introduce new variables under the integral sign by setting

$$\zeta = (\alpha - it)x, \qquad \eta = (\alpha - it)y,$$

we see easily that the integral can be written as a product of two functions, one depending on t, the other on z. Therefore, the random variables $U = X + Y$ and $V = X/Y$ are independent and the theorem is fully established.

Our theorem is closely related to a recent result of R. G. Laha [1], who characterized the gamma distribution essentially by the independence of the ratio of certain quadratic forms from the mean. Laha assumed that the random variables are identically distributed and that their second moment exists; both these assumptions were avoided in the present paper.

REFERENCES

[1] R. G. LAHA, "On a characterization of the gamma distribution," *Ann. Math. Stat.*, Vol. 25(1954), pp. 784–787.

[2] E. J. G. PITMAN, "The closest estimates of statistical parameters," *Proc. Cambridge Philos. Soc.*, Vol. 33 (1937), pp. 212–222.

THE RATIO OF VARIANCES IN A VARIANCE COMPONENTS MODEL[1]

By W. A. Thompson, Jr.

Virginia Polytechnic Institute and *University of North Carolina*

Summary. Our discussion will concern primarily λ, the ratio of two variances which arise in discussing the "mixed" incomplete block model. In Section 1 we find first the class of invariant statistics for a test involving this ratio and second the joint distribution of these statistics. In Section 2 we use these statistics to construct a test (with certain optimum properties) of the hypothesis $\lambda < \lambda_0$ versus $\lambda > \lambda_1$.

1. The general incomplete block variance components model. Suppose that y_{ij} for $i = 1, \cdots, u$, and $j = 1, \cdots, b$ are independent and normal for given $t_1, \cdots, t_u$ with means $E(y_{ij} \mid t) = n_{ij}(t_i + b_j)$ and variance σ^2. Here n_{ij} is 1 or 0 according as the ith treatment does or does not occur in the jth block. The total number of observations is N, that is, $\sum_{i,j} n_{ij} = N$. In addition suppose that the t's are independent and identically normal with mean 0 and variance ϵ^2. If t were an unknown parameter instead of a random variable, we would have the general incomplete block model which appears in analysis of variance (see, for example, Bose [1]).

In the general theory of incomplete block designs we make use of the block totals $B_1, \cdots, B_b$ and of the "adjusted yields" $Q_1, \cdots, Q_u$. It is known from this theory that the latter form the basis of a vector space V_B of dimensionality b, while the former generate a vector space V_Q of dimensionality r, say. (In the case of a connected design, $r = u - 1$.) Further V_B and V_Q are orthogonal to each other and to the error space V' of dimensionality $N - b - r$. We may now choose an orthogonal basis for V', say $Y_1, \cdots, Y_{N-b-r}$.

Again from incomplete block design theory we know that

$$\begin{aligned} E(B_j \mid t) &= k_j b_j + n_{1j}t_1 + n_{2j}t_2 + \cdots + n_{uj}t_u, \\ E(Q_i \mid t) &= c_{i1}t_1 + c_{i2}t_2 + \cdots + c_{iu}t_u, \\ E(Y_i \mid t) &= 0, \end{aligned}$$

and also that the covariance matrix of the B's and Q's and Y's for fixed t is

$$\left[\begin{array}{cc|c|c} k_1 & 0 & & \\ & \ddots & 0 & 0 \\ 0 & k_b & & \\ \hline & & & \\ 0 & & C & 0 \\ & & & \\ \hline & & & \\ 0 & & 0 & I \\ & & & \end{array}\right] \sigma^2,$$

Received January 17, 1955.

[1] Extracted from a special report to the United States Air Force under Contract No. AF 18(600)-83 monitored by the Office of Scientific Research.

where $k_j = \sum_{i=1}^{u} n_{ij}$, and the C matrix, involved in both the expectations and variances of Q, is again from incomplete block design theory.

Now we state several lemmas which were independently developed by Madow [5] and Skibinsky [6] and which we will find useful.

LEMMA 1. $E[E(X \mid Z)] = E(X)$.

LEMMA 2. $\operatorname{Var}(X) = E[\operatorname{Var}(X \mid Z)] + \operatorname{Var}[E(X \mid Z)]$.

LEMMA 3. $\operatorname{Cov}(X, Y) = E[\operatorname{Cov}(X, Y \mid Z)] + \operatorname{Cov}[E(X \mid Z), E(Y \mid Z)]$.

Applying these lemmas, we find the unconditional means and the covariance matrix to be

$$E(B_j) = k_j b_j, \qquad E(Q_i) = E(Y_i) = 0,$$

$$\operatorname{Cov}(B, Q, Y) = \left[\begin{array}{c|c|c} \begin{matrix} k_1 & & 0 \\ & \ddots & \\ 0 & & k_b \end{matrix} & 0 & 0 \\ \hline 0 & C & 0 \\ \hline 0 & 0 & I \end{array}\right] \sigma^2 + \left[\begin{array}{c|c|c} * & * & 0 \\ \hline * & C^2 & 0 \\ \hline 0 & 0 & 0 \end{array}\right] \epsilon^2,$$

where the partitions with the stars in them are known though perhaps complicated constants.

Now in order to simplify the problem as far as possible we will make the transformation $Q_1, \cdots, Q_u \to Z_1, \cdots, Z_u : Q = MZ$ where M is an orthogonal matrix such that

$$M'CM = \left[\begin{array}{c|c} \begin{matrix} e_1 & & 0 \\ & \ddots & \\ 0 & & e_r \end{matrix} & 0 \\ \hline 0 & 0 \end{array}\right] = \left[\begin{array}{c|c} D_e & 0 \\ \hline 0 & 0 \end{array}\right],$$

a diagonal matrix with the characteristic roots of C in the diagonal. Note also that

$$M'C^2M = M'C(MM')CM = (M'CM)\,(M'CM)$$

$$= \left[\begin{array}{c|c} D_{e^2} & 0 \\ \hline 0 & 0 \end{array}\right].$$

Thus $Z_1, \cdots, Z_r$ have the covariance matrix $D_e\sigma^2 + D_{e^2}\epsilon^2$ while $Z_{r+1}, \cdots, Z_n$ are zero with probability one.

Because of the orthogonality of the Z's among themselves and the mutual orthogonality of V_B, V_Q, and V', the Y's, B's, and $Z_1, \cdots, Z_r$ are N linearly independent linear functions in the space of the y's and thus form a basis for the y space. We may therefore make a transformation of the y's into the Y's, B's, and $Z_1, \cdots, Z_r$. These last variables are of course multivariate normal. From this and the nature of the covariance matrix of these final variates we see that $B_1, \cdots, B_b, Z_1, \cdots, Z_r$ and $\sum Y_i^2$ are a set of sufficient statistics for the distribution.

Suppose now that we are interested in placing confidence limits on, or testing hypothesis concerning, the ratio $\epsilon^2/\sigma^2 = \lambda$, say. Let us consider a group G of transformations on our set of sufficient statistics. Let G be

$$B_j' = cB_j + k_jc_j, \qquad Z_1' = cZ_1, \cdots, Z_r' = cZ_r, \qquad (\sum Y_i^2)' = c^2(\sum Y_i^2).$$

Since the effect of G is only to change the mean of B_j and multiply the covariance matrix of (B, Z, Y) by an arbitrary constant, c^2, the problem is invariant under G. In this connection, see Lehmann [4]. A maximal invariant under G is

$$Z_1 / \sqrt{\sum Y_i^2}, \quad \cdots, \quad Z_r / \sqrt{\sum Y_i^2}.$$

Thus G induces the group of transformations $\bar{G}$,

$$b_j' = c(b_j + c_j), \qquad \sigma^{2\prime} = c^2\sigma^2, \qquad \epsilon^{2\prime} = c^2\epsilon^2,$$

a maximal invariant for which is $\epsilon^2/\sigma^2 = \lambda$. Thus if we adhere to the principle of invariance, then in making inferences about λ, we may restrict ourselves to functions of

$$Z_1 / \sqrt{\sum Y_i^2}, \quad \cdots, \quad Z_r / \sqrt{\sum Y_i^2}.$$

We now find the joint distribution of the statistics

$$X_1 = Z_1 / \sqrt{e_1 X_{r+1}}, \quad \cdots, \quad X_r = Z_r / \sqrt{e_r X_{r+1}},$$

where $X_{r+1} = \sum_1^n Y_i^2$ and $n = N - b - r$.

Let $W_i = Z_i / \sqrt{e_i}$; then W_i is $N(0, \sigma^2 + e_i\epsilon^2)$ and since the W's and X_{r+1} are independent, their joint frequency function is

$$\text{const}(x_{r+1})^{n/2-1} \exp\left[-\frac{1}{2}\left(\frac{w_i^2}{\sigma^2 + e_i\epsilon^2} + \frac{x_{r+1}}{\sigma^2}\right)\right].$$

Making the transformation

$$X_1 = W_1 / \sqrt{X_{r+1}}, \quad \cdots, \quad X_r = W_r / \sqrt{X_{r+1}},$$

we find that the probability element of $X_1, X_2, \cdots, X_r, X_{r+1}$ is

$$\text{const}(x_{r+1})^{(n+r)/2-1} \exp\left[-\frac{x_{r+1}}{2\sigma^2}\left(1 + \sum_{i=1}^r \frac{x_i^2}{1 + e_i\lambda}\right)\right].$$

We may now integrate out over X_{r+1}, noting that we have a gamma function in this variable. We find the probability element of $X_1, \cdots, X_r$ to be

$$\text{const}\left(1 + \sum_{i=1}^{r} \frac{x_i^2}{1 + e_i\lambda}\right)^{-(n+r)/2}.$$

2. A one-sided test on λ. Let ω_0: $\lambda \leqq \lambda_0$ and ω_1: $\lambda \geqq \lambda_1$ where $\lambda_0 < \lambda_1$. We now interest ourselves in tests of $H_0: \lambda \,\varepsilon\, \omega_0$ versus $H_1: \lambda \,\varepsilon\, \omega_1$. The region between λ_0 and λ_1 is a zone of indifference to be determined by the experimental situation. That is, if $\lambda_0 < \lambda < \lambda_1$, then we do not particularly care whether we accept H_0 or H_1.

Now consider an a priori distribution defined on ω_0 which assigns probability 1 to $\lambda = \lambda_0$, and similarly a distribution on ω_1 which assigns probability 1 to $\lambda = \lambda_1$.

According to a theorem of Lehmann [3], if now we construct a most powerful size α test of λ_0 versus λ_1 which has power β and if we can show that this test has size α for the composite hypothesis and power $\geqq \beta$ for all λ in ω_1, then this test is the one which maximizes the minimum power.

We may use the Neyman-Pearson Lemma to construct a most powerful test of $\lambda = \lambda_0$ versus $\lambda = \lambda_1$. If we let

$$R = \frac{1 + \sum X_i^2 / (1 + e_i\lambda_0)}{1 + \sum X_i^2 / (1 + e_i\lambda_1)} = \frac{\sum Y_i^2 + \sum Z_i^2 / (e_i + e_i^2\lambda_0)}{\sum Y_i^2 + \sum Z_i^2 / (e_i + e_i^2\lambda_1)},$$

then the above test becomes

$$\text{if } R > c, \text{ accept hypothesis } \lambda = \lambda_1,$$
$$\text{if } R < c, \text{ accept hypothesis } \lambda = \lambda_0.$$

Here c is a constant chosen so that the test has significance level α.

The power function of this test is

$$B(\lambda) = \text{const} \int_{R>c} \exp\left[-\frac{1}{2}\left(\frac{\sum y_i^2}{\sigma^2} + \sum \frac{z_i^2}{e_i\sigma^2 + e_i^2\epsilon^2}\right)\right] dy\, dz$$
$$= \text{const} \int_{R'(\lambda)>c} \exp\left[-\frac{1}{2}\left(\sum f_i^2 + \sum g_i^2\right)\right] df\, dg,$$

where we have made the transformation

$$F_i = Y_i/\sigma, \qquad i = 1, \ldots, n;$$
$$G_i = Z_i/(e_i\sigma^2 + e_i^2\epsilon^2)^{1/2}, \qquad i = 1, \cdots, r;$$
$$R'(\lambda) = \frac{\sum F_i^2 + \sum G_i^2(1 + e_i\lambda) / (1 + e_i\lambda_0)}{\sum F_i^2 + \sum G_i^2(1 + e_i\lambda) / (1 + e_i\lambda_1)}.$$

We may compute, by straightforward though lengthy algebra, that

$$\frac{\partial R'}{\partial \lambda} = \frac{(\sum F_i^2)(\sum e_i G_i^2 / (1 + e_i\lambda_0) - \sum e_i G_i^2 / (1 + e_i\lambda_1))}{(\sum F_i^2 + \sum G_i^2(1 + e_i\lambda) / (1 + e_i\lambda_1))^2}$$
$$+ \frac{\sum_{j>i} G_i^2 G_j^2((e_j - e_i)^2(\lambda_1 - \lambda_0) / (1 + e_i\lambda_1)(1 + e_j\lambda_0)(1 + e_j\lambda_1)(1 + e_i\lambda_0)}{(\sum F_i^2 + \sum G_i^2(1 + e_i\lambda) / (1 + e_i\lambda_1))^2}$$

which is greater than 0 except when

$$F_1 = F_2 = \cdots = F_n = G_1 = G_2 = \cdots = G_r = 0.$$

Thus R' is an increasing function of λ. Also if $R'_1 \leqq R'_2$, then $c < R'_1$ implies that $c < R'_2$, so that $\int_{R'_1 > c} \leqq \int_{R'_2 > c}$. Therefore β is an increasing function of R' and thus of λ. Thus

$$\beta(\lambda) \leqq \beta(\lambda_0) = \alpha \text{ for all } \lambda \in \omega_0, \quad \beta(\lambda_1) \leqq \beta(\lambda) \text{ for all } \lambda \in \omega_1.$$

Accordingly we have proved the

THEOREM. *The test, accept or reject H_0 according as $R < c$ or $R > c$, is the one which maximizes the minimum power among all invariant tests.*

REFERENCES

[1] R. C. BOSE, "Least square aspects of analysis of variance", Institute of Statistics, University of North Carolina.

[2] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[3] E. L. LEHMANN, *Theory of Testing Hypotheses*, Associated Students Store, University of California, chapter 6, p. 10.

[4] E. L. LEHMANN, "Some principles of the theory of testing hypothesis", *Ann. Math. Stat.*, Vol. 21 (1950), pp. 8–11.

[5] WILLIAM G. MADOW, "On the theory of systematic sampling, II", *Ann. Math. Stat.*, Vol. 20 (1949), pp. 333–354.

[6] MORRIS SKIBINSKY, "A statistical model and test for the detection of gonial crossing over", Unpublished thesis, University of North Carolina.

ON MOMENTS OF ORDER STATISTICS FROM THE WEIBULL DISTRIBUTION

By Julius Lieblein

National Bureau of Standards

Summary. This note expresses the first two moments of the order statistics in samples from the Weibull distribution (sometimes referred to as the "third" asymptotic distribution of extreme values) in terms of known (incomplete B and Γ) functions. A similar procedure is applied to the "second" asymptotic distribution of extreme values.

1. Introduction. Explicit formulas in terms of a certain tabulated function were derived in an earlier paper [1] for the moments of order statistics for the "first" asymptotic distribution of largest values, with cdf

$$F(x) = \exp(-e^{-y}), \qquad y = (x - u)/\beta, \qquad -\infty < x < \infty. \tag{1.1}$$

This note extends the procedure to the other two asymptotic forms. However, the distributions for *smallest* rather than largest values are considered, with a view to possible application to breaking strength and fatigue problems. Without loss in generality for our purposes, the two other distributions may be taken (see [3]) as having cdf's

$$G(x) = \begin{cases} 1 - \exp[-(-x)^{-m}], & x \leqq 0, \\ 1, & x > 0; \end{cases} \tag{1.2}$$

$$H(x) = \begin{cases} 0, & x \leqq 0, \\ 1 - \exp(-x^{m}), & x > 0. \end{cases} \tag{1.3}$$

The parameter m is positive. These distributions are designated, respectively, the second and third asymptotic types (for *smallest* values). The distribution (1.3) has been applied extensively, mainly by Weibull [5], [6], [7], [8] and will be referred to by his name.

2. General approach. The main point of difficulty is in evaluating the double integrals that occur in the covariances, the means presenting little difficulty. It is interesting to proceed generally and try to discover other distributions for which the technique introduced in [1] will work.

If $P(x)$ is the cdf of the parent population, and $p(x)$ is the corresponding df, then the joint df of the ith and jth order statistics from $P(x)$ is [9]

$$p(x, y) = K[P(x)]^{i-1}[P(y) - P(x)]^{j-i-1}[1 - P(y)]^{n-j}p(x)p(y), \qquad -\infty < x \leqq y < \infty, \tag{2.1}$$

Received November 12, 1954.

where $K = K(n, i, j) = n!/[(i-1)!\,(j-i-1)!\,(n-j)!]$ and $x = x_i$ and $y = x_j$, with $i < j$.

If in (2.1) we write $P(y) - P(x) = [1 - P(x)] - [1 - P(y)]$, the covariances may be expressed in terms of integrals of the forms

$$
\begin{aligned}
&\int_{-\infty}^{\infty}\int_{-\infty}^{y} xy\,[P(x)]^t\,[1 - P(y)]^u\,p(x)\,p(y)\,dx\,dy,\\
(2.2)\qquad &\\
&\int_{-\infty}^{\infty}\int_{-\infty}^{y} xy\,[1 - P(x)]^t\,[P(y)]^u\,p(x)\,p(y)\,dx\,dy.
\end{aligned}
$$

Leaving aside simple cases such as the rectangular distribution $P(x) = x$, parabolic distributions $P(x) = x^k$, and perhaps other simple forms for which the evaluation procedure in [1] is unnecessary, it will be seen on trial that success in using this procedure depends upon having $P(x)$ or $1 - P(x)$ of the form $e^{Q(x)}$, where Q is a reasonably simple function. Since $P(x)$ and $1 - P(x)$ are non-negative, this is always possible with real $Q(x)$.

The integral (2.2) is then a (finite) linear combination of integrals of the form

$$\psi(t, u) = \int_{-\infty}^{\infty}\int_{-\infty}^{y} xye^{tQ(x)+uQ(y)}\,Q'(x)\,Q'(y)\,dx\,dy.$$

Here t and u are not necessarily the same as in (2.2).

To determine Q, we carry out the procedure as far as possible. The first step, integration by parts, gives

$$\psi(t, u) = \frac{1}{t}\int_{-\infty}^{\infty} ye^{(t+u)\,Q(y)}\,Q'(y)\,dy - \frac{1}{t}\int_{-\infty}^{\infty}\int_{-\infty}^{y} ye^{tQ(x)+uQ(y)}\,Q'(y)\,dx\,dy.$$

Calling the double integral $\psi_1(t, u)$, the second step, differentiating with respect to t, gives

$$\frac{\partial\psi_1(t, u)}{\partial t} = \int_{-\infty}^{\infty}\int_{-\infty}^{y} y\,Q(x)\,Q'(y)\,e^{tQ(x)+uQ(y)}\,dx\,dy.$$

This can be evaluated in two cases, (i) when the inner integral can be evaluated as it stands, and (ii) when the double integral is, apart from a constant factor, precisely of the form $\psi(t, u)$.

In case (i), we need to have $Q(x) = cQ'(x)$, whence $Q(x) = ae^{bx}$, which makes either $P(x)$ or $1 - P(x)$ doubly exponential. That is, we have the first asymptotic form for smallest values, $1 - \exp(-e^x)$. The simple transformation $x = -x'$ converts this into the distribution of largest values, $P(x) = \exp(-e^{-x})$, previously considered [1]. Thus this case has, essentially, already been treated.

Case (ii) requires that $xQ'(x) = cQ(x)$, so that $Q(x) = ax^b$ (which, of course, includes the exponential distribution). This gives, essentially, just the two forms (1.2) and (1.3).

It is rather interesting that the three asymptotic distributions of extreme

values thus seemingly exhaust the possibilities for the method introduced in connection with one of them.

3. Results. The Weibull distribution (1.3) is of chief interest. The method in case (ii) applied to (2.1) and (2.2) for this distribution yields a simple first order differential equation in $\psi(t, u)$ whose solution is given by

$$t^{1+1/m}\psi(t, u) - u^{1+1/m}\psi(u, u) = \frac{1}{m^2}\Gamma(2 + 2/m)\int_u^t t^{1/m}(t+u)^{-2-2/m}\,dt.$$

The function $\psi(u, u)$ is readily evaluated and the integral may be expressed in terms of the B-function by the change of variable $t = uw / (1 - w)$. The final expression for $\psi(t, u)$ is

$$\psi(t, u) = m^{-2}(tu)^{-r}\,\Gamma(2r)\,B_p(r, r), \qquad r = 1 + 1/m; \quad p = t/(t + u); \quad m, t, u > 0. \tag{3.1}$$

For given n, the values of ψ are needed only for $t + u \leqq n$. To cut the calculations in half, we can use the relation

$$\begin{aligned}\psi(t, u) + \psi(u, t) &= m^{-2}(tu)^{-r}\,\Gamma(2r)\,[B_p(r, r) + B_{1-p}(r, r)]\\ &= m^{-2}(tu)^{-r}\Gamma^2(r).\end{aligned}$$

The above values enable us to write down the results for the Weibull distribution (1.3) as

$$\begin{aligned}E(x_i^k) &= \frac{n!}{(i-1)!\,(n-i)!}\,\Gamma\left(1+\frac{k}{m}\right)\sum_{\mu=0}^{i-1}(-1)^\mu C_\mu^{i-1}\\ &\qquad\cdot(n+\mu-i+1)^{-1-k/m},\\ &\qquad m > 0;\quad i = 1, 2, \cdots, n;\quad k = 0, 1, 2, \cdots;\\ E(x_ix_j) &= m^2K\sum_{\mu=0}^{i-1}\sum_{\nu=0}^{j-i-1}(-1)^{\mu+\nu}C_\mu^{i-1}C_\nu^{j-i-1}\psi(j-i+\mu-\nu,\\ &\qquad n-j+\nu+1),\\ &\qquad m > 0;\quad i < j;\quad i, j = 1, 2, \cdots, n.\end{aligned} \tag{3.2}$$

Here K is as in (2.1), and the function ψ is given by (3.1).

For the second asymptotic distribution (1.2), the results are similar and may be obtained with little difficulty. The function corresponding to ψ is

$$\psi^*(t, u) = m^{-2}(tu)^{-r'}[\Gamma^2(r') - \Gamma(2r')B_p(r', r')], \quad r' = 1 - 1/m, \qquad p = t/(t + u), \quad m > 2, \quad t,u > 0. \tag{3.3}$$

The resulting moments are then

$$E(x_i^k) = \frac{n!}{(i-1)!\,(n-i)!}\Gamma\left(1 - \frac{k}{m}\right)\sum_{\mu=0}^{i-1}(-1)^{\mu+k}C_\mu^{i-1} \cdot (n + \mu - i + 1)^{-1+k/m}$$

$$m > 0;\quad i = 1, 2, \cdots, n \quad k < m;\quad k = 0, 1, 2, \cdots;$$

(3.4)

$$E(x_i x_j) = m^2 K \sum_{\mu=0}^{i-1}\sum_{\nu=0}^{j-i-1}(-1)^{\mu-\nu}\, C_\mu^{i-1} C_\nu^{j-i-1} \cdot \psi^*(n - j + \nu + 1,\; j - i + \mu - \nu),$$

$$m > 2;\quad i < j;\quad i, j, = 1, 2, \cdots, n.$$

Here K is as in (2.1), and ψ^* is given by (3.3).

4. Remarks on application. The present results can be used in cases where m is known, perhaps from previous work. One important application of moments of order statistics is in finding minimum-variance unbiased estimators by means of linear functions of such statistics. An illustration of this for the case of the first asymptotic distribution will be found in [2], Appendix C. If m is not known, then there is available [4] a simple graphical procedure for obtaining an estimate which could then be used in the above formulas for first approximations.

REFERENCES

[1] J. Lieblein, "On the exact evaluation of the variances and covariances of order statistics in samples from the extreme-value distribution," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 282–287.

[2] J. Lieblein, *A New Method of Analyzing Extreme-value Data*, NACA Tech. Note, no. 3053, January 1954.

[3] *Probability Tables for the Analysis of Extreme-value Data*, National Bureau of Standards Applied Mathematics Series, no. 22 (1953), U. S. Government Printing Office, Washington, D. C., p. 2.

[4] H. C. S. Thom, "Frequency of maximum wind speeds," *Proc. Amer. Soc. Civil Engrs.*, Vol. 80 (1954), separate no. 539, 11 pp.

[5] W. Weibull, "A statistical theory of the strength of materials," *Ing. Vetenskaps Akad. Handl.*, Vol. 151 (Stockholm 1939), p. 15.

[6] W. Weibull, "The phenomenon of rupture in solids," *Ing. Vetenskaps Akad. Handl.*, Vol. 153 (Stockholm 1939), p. 17.

[7] W. Weibull, "A statistical distribution of wide applicability," *J. Appl. Mech.*, Vol. 18 (1951), pp. 293–297.

[8] W. Weibull, "Statistical design of fatigue experiments," *J. Appl. Mech.*, Vol. 19 (1952), pp. 109–113.

[9] S. S. Wilks, *Mathematical Statistics*, Princeton University Press, 1950, p. 90.

ON THE ASYMPTOTIC NORMALITY OF SOME STATISTICS USED IN NON-PARAMETRIC TESTS

By Meyer Dwass

Northwestern University

1. Summary. Let $(R_1, \cdots, R_n)$ be a random vector which takes on each of the $N!$ permutations of $(1, 2, \cdots, N)$ with equal probability, $1/N!$. Let $(a_{N1}, \cdots, a_{NN})$ and $(b_{N1}, \cdots, b_{NN})$ be two sets of real numbers given for every N. We will assume throughout that for no N are the a_{Ni} all equal or the b_{Ni} all equal. We also assume that the a_{Ni} and b_{Ni} have been so normalized that

$$\sum a_{Ni} = \sum b_{Ni} = 0; \qquad \sum a_{Ni}^2 = N^{-1} \sum b_{Ni}^2 = 1.$$

Unless otherwise stated, $\sum$ will mean $\sum_{i=1}^{N}$. Define

$$S_N = \sum a_{Ni} b_{NR_i}. \tag{1.1}$$

Let $\Phi(x)$ be the unit normal c.d.f. In Section 2 sufficient conditions are given for $\Pr\{S_N < x\}$ to approach $\Phi(x)$ as $N \to \infty$. (The first two moments of S_N are 0 and $N/(N-1)$, respectively.)

For every N, let $Y = (Y_{11}, \cdots, Y_{1N_1}, \cdots, Y_{m1}, \cdots, Y_{mN_m})$ be $N = N_1 + \cdots + N_m$ random variables which are mutually independent and independent of the R_i. We assume that all Y_{ij} with the same first subscript are identically distributed. Define

$$\bar{Y} = N^{-1} \textstyle\sum_{i,j} Y_{ij}, \qquad S^2 = N^{-1} \sum_{i,j} (Y_{ij} - \bar{Y})^2, \qquad SY'_{ij} = Y_{ij} - \bar{Y},$$
$$Y'_{ij} = 0 \text{ if } S = 0.$$

Let Y' denote the vector of the Y'_{ij}. Let $y = (y_1, \cdots, y_N)$ denote a point in N-space. By $F_N(x, y)$ we mean the c.d.f. of the random variable $S_N = \sum a_{Ni} y_{NR_i}$. In Section 3 are considered sufficient conditions for convergence with probability one of the random c.d.f. $F_N(x, Y')$ to $\Phi(x)$.

2. Asymptotic distribution of S_N when the b_{Ni} are nonrandom. Let $G_N(x)$ denote the c.d.f. of the b_{Ni} (continuous to the left). We assume that the b_{Ni} have been so indexed that $b_{N1} \leqq b_{N2} \leqq \cdots \leqq b_{NN}$.

Theorem 2.1. *Suppose*

A) *there is a* c.d.f. $G(x)$ *such that* $\lim_{N\to\infty} G_N(x) = G(x)$ *at every point of continuity of* $G(x)$ *and*

$$\int_{-\infty}^{\infty} x \, dG(x) = 0, \qquad \int_{-\infty}^{\infty} x^2 \, dG(x) = 1;$$

and either

B) $\lim_{N\to\infty} \max_{1 \leq i \leq N} | a_{Ni} | = 0$ or B') $G(x) = \Phi(x)$.

Received July 13, 1953, revised July 21, 1954.

Then S_N (1.1) *is asymptotically normally distributed. That is,* $\Pr\{S_N < x\} \to \Phi(x)$ *as* $N \to \infty$.

This theorem can be compared with that of a previous paper [1]. There it was assumed that the b_{Ni} were essentially powers of expected values of order statistics. In this paper we use the fact that the b_{Ni} behave asymptotically like expected values of order statistics. The idea of the earlier theorem [1] is then used in proving this theorem. In [1] the counterpart of $G(x)$ was assumed continuous; in this paper we make no such assumption. Theorem 3.1, below, deals with the case where the b_{Ni} are random variables. From now on, the function $G(x)$ will be assumed to have the properties stated in Theorem 2.1.

The proof of this theorem is aided by Lemmas 2.1 to 2.7 below. The main lines of the proof are as follows. Let $X_1, \cdots, X_N$ be independent, identically distributed random variables, each with the c.d.f. $G(x)$; let $Z_{N1} \leqq \cdots \leqq Z_{NN}$ be the ordered values of the X's. We will define below a vector function of the X's, $(R_1, \cdots, R_N)$, which has the property that it assumes each of the permutations of $(1, \cdots, N)$ with equal probability, $1/N!$. Then $\sum a_{Ni} b_{NR_i} = S_N$ is the random variable whose asymptotic normality we seek to prove.

Either condition B) or B′) will assure us of the asymptotic normality of $\sum a_{Ni}X_i$. (See [1]; however, existence of third moment is not required. This follows from the counterpart of Lindeberg's condition for a sequence of sequences of random variables. I am obliged to Prof. W. Hoeffding for pointing this out to me.)

Hence it is sufficient to show that $\sum a_{Ni}X_i - S_N$ converges in probability to zero. Sufficient for this is to show that $\lim_{N\to\infty} E(\sum a_{Ni}X_i - S_N)^2 = 0$. We have that

$$E(\textstyle\sum a_{Ni}X_i - S_N)^2 = EX_1^2 - 2E(\sum a_{Ni}X_i)(S_N) + N/(N-1). \tag{2.1}$$

The purpose of Lemma 2.1 is to show that our particular definition of $(R_1, \cdots, R_N)$ provides us with the fact that $E(\sum a_{Ni}X_i)(S_N) = (N-1)^{-1} \cdot \sum EZ_{Ni}b_{Ni}$. Since $EX_1^2 = 1$, then if $\lim_{N\to\infty}(N-1)^{-1}\sum EZ_{Ni}b_{Ni} = 1$, this will imply that (2.1) approaches zero as $N \to \infty$. Making use of the fact [3] that $\lim_{N\to\infty} N^{-1}\sum(EZ_{Ni})^2 = 1$, it is easy to see that

$$\lim_{N\to\infty} N^{-1}\textstyle\sum EZ_{Ni}b_{Ni} = 1 \qquad \text{is equivalent to} \qquad \lim_{N\to\infty} N^{-1}\sum(b_{Ni} - EZ_{Ni})^2 = 0.$$

The truth of this last limit is shown in the sequence of Lemmas 2.2 through 2.7, where important use is made of some results of Hoeffding [3].

A random vector $R = (R_1, \cdots, R_N)$ is defined by considering the ordered arrangement of the X's,

$$X_{i_1} \leqq X_{i_2} \leqq \cdots \leqq X_{i_N}. \tag{2.2}$$

If no two of the X's are equal to each other, then let $R_{i_1} = 1$, $R_{i_2} = 2, \cdots, R_{i_N} = N$. If there are ties among the X's then let

$$(X_{i_1} \leqq X_{i_2} \leqq \cdots \leqq X_{i_N}), \quad \cdots, \quad (X_{j_1} \leqq X_{j_2} \leqq \cdots \leqq X_{j_N}),$$

be the set of all possible distinct ordered arrangements of the X's. Say there are p such arrangements. In such a case, let

$$P\{(R_{i_1} = 1),\ (R_{i_2} = 2),\ \cdots,\ (R_{i_N} = N) \mid (X_1, \cdots, X_N)\} = 1/p,$$
$$\cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots$$
$$P\{(R_{j_1} = 1),\ (R_{j_2} = 2),\ \cdots,\ (R_{j_N} = N) \mid (X_1, \cdots, X_N)\} = 1/p.$$

It is easy to see that R equals each of the $N!$ permutations of $(1, \cdots, N)$ with probability $1/N!$

LEMMA 2.1. $E(\sum a_{Ni}X_i)\ (\sum a_{Ni} b_{NR_i}) = (N-1)^{-1} \sum EZ_{Ni} b_{Ni}$.

PROOF. The left side can be written as

$$(N!)^{-1} \sum{}' \{\sum a_{Ni} b_{Nr_i} E(\sum a_{Ni} X_i \mid (R_1 = r_1), \cdots, (R_N = r_N)\}$$

where $(r_1, \cdots, r_N)$ is one of the permutations of $(1, \cdots, N)$ and $\sum'$ is summation over all $N!$ permutations. We also have that

$$E[X_1 \mid (R_1 = r_1), \cdots, (R_N = r_N)] = E[Z_{Nr_i} \mid (R_1 = r_1), \cdots, (R_N = r_N)].$$

Since the distribution of $X_1, \cdots, X_N$ is invariant under all permutations, this last conditional expectation does not depend on the values of the R's and hence equals EZ_{Nr_i}. This shows that the left side equals

$$(N!)^{-1} \sum{}' (\sum a_{Ni} b_{Nr_i})\ (\sum a_{Ni} EZ_{Nr_i}).$$

By an elementary calculation, this in turn is equal to the right side.

Since $Z_{N1} \leqq Z_{N2} \leqq \cdots \leqq Z_{NN}$, then $EZ_{N1} \leqq EZ_{N2} \leqq \cdots \leqq EZ_{NN}$. Denote by $H_N(x)$ the c.d.f. of the EZ_{Ni} (continuous to the left).

LEMMA 2.2 (Hoeffding). *$\lim_{N\to\infty} H_N(x) = G(x)$ for every x which is a point of continuity of $G(x)$.*

This is proved by Hoeffding ([3], Theorems 1 and 2).

Let S be that set of points on the real line where $G(x)$ is discontinuous, together with all points x where $G(x - h) < G(x) < G(x + h)$ for every $h > 0$.

LEMMA 2.3. *Given $\epsilon_1 > 0$ and $\epsilon_2 > 0$, there is an even integer $2n$ and there is a sequence of closed intervals which are mutually disjoint,*

$$[t_1, t_2], [t_3, t_4], \cdots, [t_{2n-1}, t_{2n}],$$

having the following properties.

a) *With each t_i we can associate an a_i with $0 < a_i < 1$ and a $y_i \in S$ such that*

$$t_i = a_i\, G(y_i - 0) + (1 - a_i)\, G(y_i + 0);$$

b) $y_1^2(t_2 - t_1) + y_3^2(t_4 - t_3) + \cdots + y_{2n-1}^2(t_{2n} - t_{2n-1}) > 1 - \epsilon_1$;

c) $\max\,[(y_2 - y_1), (y_4 - y_3), \cdots, (y_{2n} - y_{2n-1})] < \epsilon_2$.

PROOF. We will denote the right side of the equality in part a) by $G_{a_i}(y_i)$. Since $\int_A^B y^2\, dG(y)$ exists in Riemann-Stieltjes sense for any A, B, we can find finite A', B' and a partitioning,

$$A' = y_1' < y_2' < \cdots < y_m' = B',$$

such that

$$y_1'^2[G(y_2') - G(y_1')] + y_2'^2[G(y_3') - G(y_2')] + \cdots + y_{m-1}'^2[G(y_m') - G(y_{m-1}')] > 1 - \tfrac{1}{2}\epsilon_1,$$

$$\max_{1 \le i \le m-1} (y_{i+1}' - y_i') < \tfrac{1}{2}\epsilon_2.$$

We can replace this partitioning by one

$$A = y_1 < y_2 < \cdots < y_{2n} = B,$$

and find constants $0 < a_i < 1$ where each $y_i \varepsilon S$ and for which

$$y_1^2[G_{a_2}(y_2) - G_{a_1}(y_1)] + y_3^2[G_{a_4}(y_4) - G_{a_3}(y_3)] + \cdots + y_{2n-1}^2[G_{a_{2n}}(y_{2n}) - G_{a_{2n-1}}(y_{2n-1})] > 1 - \epsilon_1,$$

and where c) is satisfied. The clumsy but elementary details are omitted. We need now only set $t_i = G_{a_i}(y_i)$ for $i = 1, 2, \cdots, 2n$; adjacent t_i's can be equal to each other.

LEMMA 2.4. *For every N, let there be given an $m = m(N)$. Let $y \varepsilon S$. If $m/N \to aG(y - 0) + (1 - a)\, G(y + 0)$, for some a where $0 < a < 1$, as $N \to \infty$, then $b_{Nm} \to y$ and $EZ_{Nm} \to y$.*

PROOF. Suppose $\overline{\lim}_{N\to\infty} b_{Nm} = y + \delta$, where $\delta > 0$. Then there are δ', where $0 < \delta' < \delta$, such that $y + \delta'$ is a continuity point of G, and a subsequence N_i, m_i, such that

$$\lim_{i\to\infty} m_i/N_i \ge G(y + \delta') > aG(y - 0) + (1 - a)\, G(y + 0),$$

which gives a contradiction. We treat $\underline{\lim}\ b_{Nm}$ similarly. Making use of Lemma 2.2, we can prove this for the EZ_{Ni} in the same way. This was proved by Hoeffding ([3], Lemma 5) without using the result quoted in our Lemma 2.2.

Let $T = [t_1, t_2] \cup [t_3, t_4] \cup \cdots \cup [t_{2n-1}, t_{2n}]$, a finite union of disjoint closed intervals. Let $\sum_T a_i$ mean summation over all indices i such that $(i/N) \varepsilon T$.

LEMMA 2.5 *For every $\epsilon > 0$, there exists a finite union of disjoint closed intervals, $[t_1, t_2] \cup \cdots \cup [t_{2n-1}, t_{2n}] = T$, such that*

$$\underline{\lim}\ N^{-1} \sum_T b_{Ni}^2 > 1 - \epsilon, \qquad \underline{\lim}\ N^{-1} \sum_T (EZ_{Ni})^2 > 1 - \epsilon.$$

PROOF. Choose $\epsilon > 0$. We can find a sequence of closed intervals which are mutually disjoint, $[t_1, t_2], [t_3, t_4], \cdots, [t_{2n-1}, t_{2n}]$, satisfying the requirements of Lemma 2.3 for $\epsilon_1 = \frac{1}{2}\epsilon$; the value of ϵ_2 of Lemma 2.3 is not material here. Then

$$N^{-1} \sum_T b_{Ni}^2 = N^{-1} \sum_{Nt_1 \le i \le Nt_2} b_{Ni}^2 + N^{-1} \sum_{Nt_3 \le i \le Nt_4} b_{Ni}^2 + \cdots + N^{-1} \sum_{Nt_{2n-1} \le i \le Nt_{2n}} b_{Ni}^2.$$

For simplicity, look at the first term on the right side of the above equality. Let p denote the number of integers contained between Nt_1 and Nt_2 inclusive. Let $[r]$ denote the least integer greater than or equal to r. Then

$$N^{-1} \sum_{Nt_1 \leq i \leq Nt_2} b_{Ni}^2 \geq b_{N[Nt_1]} N^{-1} p.$$

By Lemma 2.4 the first factor of the right side has the limit y_1^2. The second factor has the limit $t_2 - t_1$. Hence the limit inferior of the left side is greater than $y_1^2(t_2 - t_1) - \frac{1}{2}\epsilon/n$. Therefore

$$\varliminf_{N\to\infty} N^{-1} \sum_T b_{Ni}^2 > \sum y_{2j-1}^2 (t_{2j} - t_{2j-1}) - \tfrac{1}{2}\epsilon > 1 - \epsilon.$$

The last inequality of Lemma 2.5 is proved in the same way.

LEMMA 2.6. *Given $\epsilon > 0$, we can find a set $T = T(\epsilon)$, such that Lemma 2.5 holds, and also $\varlimsup_{N\to\infty} N^{-1}\sum_T (b_{Ni} - EZ_{Ni})^2 < \epsilon$.*

We omit the proof, since it is very similar to that of Lemma 2.5. It uses Lemmas 2.2–2.4; in particular, Lemma 2.3 is essential.

LEMMA 2.7. $\lim_{N\to\infty} N^{-1} \sum (b_{Ni} - EZ_{Ni})^2 = 0.$

PROOF. For $\epsilon > 0$, find $T = T(\epsilon)$ satisfying Lemma 2.6. Write

$$N^{-1} \sum (b_{Ni} - EZ_{Ni})^2 = N^{-1} \sum_T (b_{Ni} - EZ_{Ni})^2 + N^{-1} \sum{}' (b_{Ni} - EZ_{Ni})^2,$$

where $\sum'$ means summation over those indices not summed in $\sum_T$. By Lemma 2.6, for all N sufficiently large, $N^{-1}\sum_T (b_{Ni} - EZ_{Ni})^2 < 2\epsilon$. It is true that

$$N^{-1} \sum{}' (b_{Ni} - EZ_{Ni})^2 \leq \left(\sqrt{N^{-1} \sum{}' b_{Ni}^2} + \sqrt{N^{-1} \sum{}'(EZ_{Ni})^2}\right)^2.$$

For all N sufficiently large, each of the terms under the square root sign is, by Lemma 2.5, less than 2ϵ. Thus for all N sufficiently large, $N^{-1} \sum (b_{Ni} - EZ_{Ni})^2$ is arbitrarily small, which proves the lemma.

3. Asymptotic distribution of S_N when the b_{Ni} are random variables. We will now deal with the generalization of Theorem 2.1 that was described in the last paragraph of Section 1. We refer to the definitions made there. Let $G_N(x, Y)$ be the proportion of Y_{ij} which are less than x, and let $G_N'(x, Y')$ be the proportion of Y_{ij}' which are less than x. Let $G_1(x), \cdots, G_m(x)$ be the c.d.f.'s of Y_{11}, $Y_{21}, \cdots, Y_{m1}$ respectively. Let

$$\int_{-\infty}^{\infty} x\, dG_i(x) = \mu_i, \qquad \int_{-\infty}^{\infty} (x - \mu)^2\, dG_i(x) = \sigma_i^2, \qquad G(x) = \sum_{i=1}^{m} k_i G_i(x),$$

where $k_1, \cdots, k_m$ are constants defined in Theorem 3.1 below, with $k_i \geq 0$, and $\sum_1^m k_i = 1$.

THEOREM 3.1. *Let $G'(x)$ be the* c.d.f. *obtained from $G(x)$ by a linear transformation of the independent variable, so that the first two moments of $G'(x)$ are* 0 *and* 1. *Suppose*

A) $\lim_{N\to\infty} N_i/N = k_i$ *exists for $i = 1, \cdots, m$;*

B) $0 < \sigma_i^2 < \infty$;

C) *condition* B) *or* B′) *of Theorem* 2.1 *is satisfied for* $G'(x)$.

Then $F_N(x, Y')$, *as defined in Section* 1, *converges with probability one to* $\Phi(x)$.

PROOF. According to the Cantelli-Glivenko theorem, and some well-known facts about strong convergence, we have that $G_N(t, Y)$ converges with probability one to $G(t)$, uniformly in t ([2], pp. 257–279, 280). Also we have that, with probability one, $\bar{Y}$ and S^2 converge respectively to the mean and variance of a random variable distributed with c.d.f. $G(x)$. Then, with probability one, $G'_N(t, Y') \to G'(t)$ at the continuity points of $G'(t)$, as $N \to \infty$. Hence the required result follows from Theorem 2.1.

4. Applications. Some applications of a theorem like Theorem 2.1 have been pointed out previously [1]. We wish here to point out that Theorem 3.1 can be useful in evaluating the "large sample power" of tests of the sort that were studied by Hoeffding [4], specifically his Theorem 6.2. Analogous results can be obtained where the X's of that theorem are essentially like the Y's of our Theorem 3.1. Also, the results of this paper can be used to study analysis of variance tests like those of Hoeffding's Section 5. We plan a future paper on the large sample power of analysis of variance tests of this type.

REFERENCES

[1] MEYER DWASS, "On the asymptotic normality of certain rank order statistics," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 303–306.

[2] M. FRECHÉT, *Généralités sur les Probabilités*, premier livre, Gauthier-Villars, Paris, 1950.

[3] WASSILY HOEFFDING, "On the distribution of the expected values of the order statistics," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 93–100.

[4] WASSILY HOEFFDING, "The large-sample power of tests based on permutations of observations," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 169–192.

EMPIRICAL POWER FUNCTIONS FOR NONPARAMETRIC TWO-SAMPLE TESTS FOR SMALL SAMPLES[1]

BY D. TEICHROEW

University of California, Los Angeles

A number of nonparametric or "distribution-free" tests have been proposed [2] for the problem of testing the hypothesis that two samples come from the same population. Some of these tests are functions of the "ranking" sequences which are obtained by arranging the observations from both populations from smallest to largest, and then replacing the observations from the first population by zero and the observations from the second population by one.

If a sample of m is taken from the first population and a sample of n from the second, there are $(m + n)!/m!\,n!$ different sequences. (Since we are dealing with continuous distributions, the probability of a tie is zero. The computer automatically carried about eight decimal places for each normal deviate and therefore there was no need to investigate ties.) If the populations are identical, all sequences have equal probability of occurring. If the populations are not identical, some sequences have a higher probability than others.

The hypothesis can be tested by the use of a critical region. The simplest test consists of selecting, in advance, k particular rankings to be in the critical region. If one of these rankings occurs, the hypothesis will be rejected. The size of the critical region, that is, the probability of rejecting the hypothesis when it is true, is given by

$$\alpha = k \frac{m!\,n!}{(m + n)!}.$$

Non-integer values of k can be obtained by the use of randomized tests.

The problem of which rankings to place in the critical region is settled on the basis of power. Let $P(R_i)$ be the probability of a ranking under an alternative hypothesis. The power, for any alternative, is obtained by summing the $P(R_i)$ for all R_i which belong to the critical region. Therefore, to construct an optimum test, one wants a test which, as α is increased, places the rankings into the critical region in the order of decreasing $P(R_i)$.

Clearly, this test is the optimum test against the alternative, since it maximizes the power. In general it has been possible to construct such optimum tests only against simple alternatives. It is of some interest to determine whether the construction yields optimum tests against classes of alternatives, that is, whether uniformly most powerful tests exist.

Consider the special problem of testing the hypothesis that two samples come

---

Received November 12, 1954, revised December 15, 1954.

[1] The preparation of this paper was sponsored (in part) by the Office of Naval Research, USN.

TABLE I

$m = 3$		δ								
Ranking	c_1	0.00	.25	.50	.75	1.00	1.25	1.50	2.00	2.50
					$n = 2$					
00011	exact	10.00	6.44	3.94	2.29	1.3	.65	.32	.1	
00011	−1.66	9.0	6.50	4.10	2.25	1.3	.20	.10		
00101	−1.16	11.0	7.30	5.10	3.45	1.1	1.40	.85	.2	
01001	−.66	11.3	8.00	6.30	4.45	2.9	1.95	1.45	.2	
00110	−.50	9.8	8.65	6.65	5.45	4.6	2.30	1.35	.5	
01010	.00	9.3	9.15	8.70	7.40	5.8	4.55	3.20	1.4	
10001	.00	10.0	10.60	9.85	8.10	6.6	5.35	3.70	1.5	
01100	.50	10.3	11.55	11.25	11.15	11.2	8.70	7.20	3.7	
10010	.66	9.0	11.35	12.25	12.00	12.6	11.05	9.20	4.4	
10100	1.16	9.9	12.40	15.00	18.45	19.7	20.90	20.50	18.6	
11000	1.66	10.4	14.50	20.80	27.30	34.2	43.60	52.45	69.5	
11000	exact	10.00	14.77	20.81	28.05	36.24	45.04	54.01	70.72	
					$n = 3$					
000111	exact	5.00	2.859	1.533	.770	.361	.157	.064	.008	.001
000111	−2.11	4.90	2.825	1.675	.875	.400	.175	.075	.025	
001011	−1.71	4.60	2.950	1.875	.750	.400	.275	.100	.025	
010011	−1.27	5.00	3.175	1.825	1.250	.800	.375	.100	.025	.05
001101	−1.27	4.30	3.550	2.075	1.475	.925	.375	.200	.025	
010101	−.83	5.60	3.700	2.375	1.825	1.025	.600	.350	.175	
100011	−.64	5.00	4.275	3.250	2.325	1.750	.800	.475	.175	.05
001110	−.64	5.35	4.275	3.600	2.525	1.475	.950	.550	.225	.025
011001	−.43	5.35	4.225	3.150	2.500	1.675	1.075	.600	.075	.05
010110	−.20	4.40	4.050	3.900	2.675	2.000	1.075	.750	.150	.10
100101	−.20	5.35	4.575	3.650	2.850	1.825	1.525	1.025	.125	.05
101001	.20	4.65	5.050	4.525	3.725	3.375	2.325	1.575	.500	.20
011010	.20	4.40	5.150	4.950	3.675	3.350	2.500	1.775	.675	.15
100110	.43	4.95	5.725	5.400	5.025	4.300	3.225	1.975	.825	.35
011100	.64	5.75	5.500	5.500	6.000	5.375	4.475	3.675	1.725	.70
110001	.64	5.35	5.900	6.350	6.150	4.975	4.150	3.350	1.775	.60
101010	.83	4.65	5.875	6.300	5.975	5.825	5.000	4.100	2.000	.60
101100	1.27	5.15	6.450	7.650	9.050	8.425	8.800	8.400	6.175	3.35
110010	1.27	5.05	6.550	8.300	9.175	9.750	9.850	8.600	5.550	3.15
110100	1.71	5.40	7.825	11.150	13.375	15.850	17.400	18.475	17.350	13.60
111000	2.11	4.80	8.375	12.500	18.800	26.500	35.050	43.850	62.400	76.95
111000	exact	5.00	8.222	12.748	18.697	26.025	34.499	43.721	62.357	78.116

TABLE II

$m = 4$		δ								
Ranking	c_1	0.00	.25	.50	.75	1.00	1.25	1.50	2.00	2.50
					$n = 2$					
000011	exact	6.67	4.02	2.29	1.23	.62		.13	.02	
000011	−1.91	6.28	3.52	2.02	1.26	.70		.12	.04	—
000101	−1.47	6.70	4.86	2.74	1.44	.78		.32	.02	.02
001001	−1.07	6.44	4.68	3.44	2.04	1.14		.26	.06	.06
000110	−.84	5.92	4.74	3.54	2.58	1.68		.48	.14	.02
010001	−.63	6.32	5.46	4.02	2.80	2.14		.60	.26	.04
001010	−.44	6.98	5.52	4.86	3.28	2.26		.98	.44	.04
001100	.00	7.28	6.58	5.58	4.76	3.62		1.48	.50	.20
010010	.00	7.00	6.26	5.44	4.28	3.08		1.64	.54	.24
100001	.00	6.76	6.04	5.28	4.62	3.34		2.00	.70	.18
010100	.44	6.76	7.70	7.46	6.40	5.30		3.22	1.10	.44
100010	.63	6.96	7.82	8.32	7.96	7.04		4.86	2.70	.98
011000	.84	6.64	7.88	9.26	10.04	9.74		7.24	3.90	1.78
100100	1.07	6.52	8.78	9.98	10.70	11.36		9.14	6.64	3.60
101000	1.47	6.94	9.54	13.00	15.96	18.08		19.42	16.92	11.14
110000	1.91	6.50	10.62	15.06	21.88	29.74		48.24	66.04	81.26
110000	exact	6.67	10.45	15.55	21.99	29.66		47.43	65.38	80.11
					$n = 3$					
0000111	exact	2.86	1.49	.72	.32	.13	.05	.02		
0000111	−2.46	2.77	1.50	.82	.36	.15	.06	.02	—	—
0001011	−2.11	3.34	1.68	.84	.50	.14	.08	.05	—	—
0010011	−1.76	2.68	2.06	1.07	.66	.35	.22	.07	—	—
0001101	−1.70	2.80	1.82	.88	.34	.28	.06	—	.01	—
0100011	−1.35	2.85	2.02	1.21	.66	.35	.16	.10	.01	—
0010101	−1.35	3.07	1.62	1.24	.70	.34	.10	.05	—	—
0001110	−1.11	2.78	2.06	1.35	.78	.34	.14	.05	.01	—
0011001	−1.00	2.84	2.14	1.48	.90	.62	.28	.08	.02	—
0100101	−.94	2.52	2.02	1.37	.98	.65	.22	.10	.04	—
0010110	−.76	3.32	2.58	1.68	1.18	.58	.34	.15	—	—
1000011	−.76	2.90	2.26	1.88	1.20	.55	.36	.22	.08	.05
0101001	−.59	2.70	2.04	1.64	1.06	.77	.50	.27	.01	—
0011010	−.41	2.61	2.72	1.75	1.16	.64	.52	.27	.05	—
0100110	−.35	2.80	2.34	1.98	1.38	.94	.52	.31	.05	—
1000101	−.35	2.74	2.26	1.94	1.44	.95	.56	.27	.02	.05
0110001	−.24	3.01	2.68	1.92	1.48	.97	.76	.35	.15	—
1001001	.00	3.08	2.82	2.28	2.04	1.62	.74	.62	.10	—
0101010	.00	2.57	3.06	2.37	2.02	1.50	.76	.48	.12	—
0011100	.00	3.00	2.64	2.51	2.10	1.51	1.00	.45	.12	.10
1000110	.24	2.91	2.90	2.74	2.24	1.60	1.24	.80	.20	—

TABLE II—*continued*

$m = 4$		δ								
Ranking	c_1	0.00	.25	.50	.75	1.00	1.25	1.50	2.00	2.50
					$n = 3$—*continued*					
1010001	.35	2.77	3.16	2.82	2.38	1.74	1.36	.88	.28	.05
0110010	.35	2.85	2.70	3.15	2.52	1.95	1.60	.97	.27	.10
0101100	.41	2.75	3.26	2.82	2.82	2.17	1.52	1.11	.30	—
1001010	.59	2.85	2.54	3.14	2.82	2.48	1.82	1.27	.47	.15
1100001	.76	2.57	3.16	3.58	3.70	3.15	2.66	1.71	.91	.20
0110100	.76	3.00	3.36	3.61	3.16	2.80	2.32	1.88	.77	.25
1010010	.94	3.22	3.68	3.81	4.20	3.52	3.04	2.08	1.10	.25
1001100	1.00	2.75	3.54	3.81	3.80	3.64	3.04	2.44	1.32	.25
0111000	1.11	2.61	3.46	4.15	4.44	4.32	4.14	3.61	2.20	.95
1010100	1.35	2.87	3.84	3.80	5.00	5.31	5.08	4.17	2.42	.90
1100010	1.35	2.92	4.04	4.82	5.28	5.40	5.22	4.50	2.52	1.20
1011000	1.70	2.67	4.48	5.87	6.96	7.98	8.26	7.91	6.27	4.20
1100100	1.76	3.04	4.28	5.57	6.86	8.85	9.76	9.25	6.58	4.25
1101000	2.11	2.74	3.90	7.00	9.42	12.08	14.36	18.00	17.75	13.75
1110000	2.46	2.97	5.38	8.94	13.46	19.60	27.20	35.37	55.71	73.30
1110000	exact	2.86	5.11	8.53	13.39	19.77	27.59	36.57	55.94	73.52

from the same normal population against the alternative that they come from normal populations with the same variance σ^2 but with means which differ by $\sigma\delta$. The optimum test for this problem is, of course, the t test. The optimum rank test for this problem, under the assumption that δ is small, is the c_1 test which was given by Hoeffding [1] and studied in detail by Terry [4].

An interesting practical problem arises as to the behavior of the c_1 test for large values of δ, for example, should one use the c_1 statistic in preference to the rank sum? The question may be phrased in the following way. If the rankings are ordered so that

$$P(R_1) \geqq P(R_2) \geqq P(R_3) \geqq \cdots$$

is true for small values of δ, will this condition hold for all values of δ?

Tables I and II give the empirical frequencies, in percent, of all possible rankings which are obtained when a sample of m from $N(0, 1)$ and a sample of n from $N(-\delta, 1)$ are ranked in order of size and the individual values are replaced by 0 if they come from $N(0, 1)$ and by 1 if they come from $N(-\delta, 1)$. For the two extreme rankings, exact probabilities have been computed [3] and are given for comparison. The tables also give for each ranking the corresponding c_1 value.

The frequencies were obtained as a by-product of a sampling experiment performed on the SWAC. The number of samples computed for the different values of δ varied in the experiment. The values for $n = 2$ in Table I are based on 2000

samples for $\delta = .25, .50, .75, 1.25$, and 1.50, and on 1000 samples for $\delta = 0$, 1.00, and 2.00. The values for $n = 3$ in Table I are based on 4000 samples for each δ except 2.50, in which case 2000 samples were obtained. Table II is based on 5000 samples for $n = 2$ and 7000 samples for $n = 3$, for all values of δ.

The rankings in the tables have been arranged in the order of their c_1 values; in cases where the c_1 value is the same for two or more rankings, the ordering is by increasing probability. The tables show that the probabilities increase essentially monotonically. The deviation from monotonic increase can be accounted for by sampling fluctuation. These tables, therefore, indicate that, at least for some significance levels, it may be possible to construct uniformly most powerful rank order tests for the hypothesis.

The tables may be used to estimate the power function of any rank order test for testing whether two samples from normal populations with the same variance have different means for sample sizes $(m,n) = (3,2)$, $(3,3)$, $(4,2)$, and $(4,3)$. The tests may be randomized or unrandomized and one-sided or two-sided.

I am indebted to I. R. Savage for suggesting the problem and for helpful discussions about the results.

REFERENCES

[1] W. Hoeffding, "Optimum nonparametric tests," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 83–92 (see p. 88).

[2] I. R. Savage, "Bibliography on nonparametric statistics and related topics," *J. Amer. Stat. Assn.* Vol. 48 (1953), pp. 844–906.

[3] D. Teichroew, "A table giving a probability associated with order statistics in samples from two normal populations which have the same variance but different means," in manuscript.

[4] M. E. Terry, "Some rank order tests which are most powerful against specific parametric alternatives," *Ann. Math. Stat.* Vol. 23 (1952), pp. 346–366.

NOTES

A NOTE ON THE THEORY OF UNBIASSED ESTIMATION

By D. Basu
University of California, Berkeley

1. Summary. It is shown that even in very simple situations (like estimating the mean of a normal population) where a uniformly minimum variance unbiassed estimator of the unknown population characteristic is known to exist, no best (even locally) unbiassed estimator exists as soon as we alter slightly the definition of variance.

2. Introduction. Let $(\mathfrak{X}, \mathfrak{A})$ be an arbitrary measurable space (the "sample space") and let $\{P_\theta\}$, $\theta \varepsilon \Omega$, be a family of probability measures on $\mathfrak{A}$. A real-valued function $\mu = \mu_\theta$ of θ is "estimable" if it has an "unbiassed estimator". An unbiassed estimator of μ is a mapping $\eta = \eta_x$ of the "sample space" $\mathfrak{X}$ onto the space of all probability measures over the σ-field of all the Borel sets on the real line such that

(i) $T_x = \int_{-\infty}^{\infty} t \, d\eta_x$ is an $\mathfrak{A}$-measurable function of x,

(ii) $\mu_\theta \equiv \int_{\mathfrak{X}} T_x \, dP_\theta$ for all $\theta \varepsilon \Omega$.

If, for every $x \varepsilon \mathfrak{X}$, the whole probability mass of η_x is concentrated at one point, say T_x, then η_x (or equivalently T_x) is called a nonrandomized estimator. With reference to a given loss or weight function $w(t, \theta)$, which is a Borel-measurable function of the real variable t for every fixed $\theta \varepsilon \Omega$, an unbiassed estimator η_θ of μ_θ is better than an alternative unbiassed estimator η_x' at the point $\theta = \theta_0$ if

$$\int_{\mathfrak{X}} dP_{\theta_0} \int_{-\infty}^{\infty} w(t, \theta_0) \, d\eta_x < \int_{\mathfrak{X}} dP_{\theta_0} \int_{-\infty}^{\infty} w(t, \theta_0) \, d\eta_x'.$$

We consider only such estimators η_x for which $\int_{-\infty}^{\infty} w(t, \theta) \, d\eta_x$ is an $\mathfrak{A}$-measurable function of x for all $\theta \varepsilon \Omega$.

Hodges and Lehmann [2] noted that if, for every $\theta \varepsilon \Omega$, the loss function $w(t, \theta)$ is a convex (downwards) function of the variable t, then the class of non-randomized estimators of μ is essentially complete. Barankin [1] and Stein [4] considered the particular case where $w = |t - \mu_\theta|^s$ for $s \geq 1$ and proved, under a few regularity assumptions, that there always exists an unbiassed estimator which is locally the best at a given value of $\theta = \theta_0$. Simple examples may be

Received May 12, 1954.

given to show that there need not exist a uniformly best unbiassed estimator even in the simplest case of $s = 2$. If, however, there exists a complete sufficient statistic [3] for θ and if w is convex (downwards) for every fixed $\theta \varepsilon \Omega$, then there exists an essentially unique uniformly best unbiassed estimator for every estimable parametric function μ_θ. The convexity of the loss function is essential in the proofs of the above results. We demonstrate in the next section how a slight departure from the convexity of the loss function might destroy all these results.

3. Nonexistence of a best unbiassed estimator. Let us assume that $w(t, \theta) \geqq w(\mu_\theta, \theta) = 0$ for all t and θ. That is, we assume that the loss function is nonnegative and that there is no loss when the estimate hits the mark. Let U be the class of all unbiassed estimators η_x of μ_θ for which the risk function

$$r(\theta \mid \eta) = E[w(t, \theta) \mid \eta, \theta] = \int_{\mathfrak{X}} dP_\theta \int_{-\infty}^{\infty} w(t, \theta)\, d\eta_x$$

is defined for all θ. We prove the following

THEOREM. *If for every fixed $\theta \varepsilon \Omega$ the loss function $w(t, \theta)$ is bounded in every finite interval $|t - \mu_\theta| \leqq A$, and is $o(|t - \mu_\theta|)$ as $|t - \mu_\theta| \to \infty$, then*

$$\inf_{\eta \varepsilon U} r(\theta \mid \eta) \underset{\theta}{\equiv} 0.$$

PROOF. Let $T = T_x$ be a nonrandomized unbiassed estimator of μ_θ. The existence of an unbiassed estimator clearly implies the existence of such a T_x. Consider now the randomized estimator $\eta^{(\delta)} = \eta_x^{(\delta)}$ which, for any $x \varepsilon \mathfrak{X}$, has its entire probability mass concentrated at the two points μ_{θ_0} and $(T_x - \mu_{\theta_0})/\delta + \mu_{\theta_0}$ on the real line in the ratio $1 - \delta$ to δ, with $0 < \delta < 1$. It is easily verified that $\eta^{(\delta)}$ is an unbiassed estimator of μ_θ and that

$$\begin{aligned} r(\theta_0 \mid \eta^{(\delta)}) &= E[w(t, \theta_0) \mid \eta^{(\delta)}, \theta_0] \\ &= E[\delta w(H/\delta + \mu_{\theta_0}, \theta_0) \mid \theta_0], \qquad H = T_x - \mu_{\theta_0}. \end{aligned}$$

Since $w = o(|t - \mu_{\theta_0}|)$ as $|t - \mu_{\theta_0}| \to \infty$, given $\epsilon > 0$ we can determine A so large that

$$w(t, \theta_0) \leqq \epsilon |t - \mu_{\theta_0}|, \qquad |t - \mu_{\theta_0}| \geqq A.$$

Let $B = \sup_{|t - \mu_{\theta_0}| < A} w(t, \theta_0) < \infty$. Then

$$\begin{aligned} r(\theta_0 \mid \eta^{(\delta)}) &= \left\{ \int_{|H| < \delta A} + \int_{|H| \geqq \delta A} \right\} \delta w(H/\delta + \mu_{\theta_0}, \theta_0)\, dP_{\theta_0} \\ &\leqq \delta B + \epsilon E(|H| \mid \theta_0). \end{aligned}$$

Since ϵ and δ are arbitrary and B depends only on ϵ, it follows that $\inf_{\eta \varepsilon U} r(\theta_0 \mid \eta) = 0$. Since θ_0 is arbitrary, the theorem is proved.

Now, if $w(t, \theta_0) > 0$ for $t \neq \mu_{\theta_0}$, then $r(\theta_0 \mid \eta)$ can be zero only if η_x gives

probability one to μ_{θ_0} for almost all x with respect to the measure P_{θ_0}. In the usual circumstances, η_x then would not be an unbiassed estimator of μ_θ.

Thus, this theorem shows that if we work with a loss function satisfying the conditions of the theorem, even locally best unbiassed estimators would not exist in all the familiar situations in which we are interested. In particular estimation problems, it will be easy to see that the theorem holds even in the restricted class U^* of all nonrandomized estimators of μ. In the next section we consider the classical problem of estimating the mean of a normal population, but with a slightly altered definition of variance.

4. The case of the normal mean. Let $x = (x_1, x_2, \cdots, x_n)$ be a random sample from $N(\theta, 1)$. The problem is to get a good unbiassed estimator of θ with the loss function

$$w(t, \theta) = \begin{cases} (t - \theta)^2, & |t - \theta| \leqq a, \\ a^{3/2}\, |t - \theta|^{1/2}, & |t - \theta| > a, \end{cases}$$

where a is an arbitrarily large constant.

Let $\bar{x}$ and s^2 be the sample mean and variance, respectively, and let c_δ be the upper 100δ per cent point of the probability distribution of s^2, where $0 > \delta > 1$. Consider the nonrandomized estimator

$$T^{(\delta)} = T_x^{(\delta)} = \begin{cases} \theta_0, & s^2 \leqq c_\delta, \\ (\bar{x} - \theta_0)/\delta + \theta_0, & s^2 > c_\delta. \end{cases}$$

Since the distribution of s^2 is independent of θ and $\bar{x}$, it follows that $T^{(\delta)}$ is a function of x and δ alone and that $T^{(\delta)}$, for every fixed δ with $0 < \delta < 1$, is an unbiassed estimator of θ. Also

$$\begin{aligned} r(\theta_0 \mid T^{(\delta)}) &= E[\delta w\{(\bar{x} - \theta_0) / \delta + \theta_0, \ \theta_0\} \mid \theta_0] \\ &= \int_{|\bar{x}-\theta_0| \leq a\delta} \delta \left(\frac{\bar{x} - \theta_0}{\delta}\right)^2 \phi(\bar{x})\, d\bar{x} + \int_{|\bar{x}-\theta_0| > a\delta} \delta a^{3/2} \left|\frac{\bar{x} - \theta_0}{\delta}\right|^{1/2} \phi(\bar{x})\, d\bar{x} \\ &< \delta a^2 + \delta^{1/2} a^{3/2} E(|\bar{x} - \theta_0|^{1/2} \mid \theta_0), \end{aligned}$$

where $\phi(\bar{x})$ is the frequency function of $\bar{x}$ when $\theta = \theta_0$. Thus $r(\theta_0 \mid T^{(\delta)}) \to 0$ as $\delta \to 0$. Therefore

$$\inf_{T \in U^*} r(\theta \mid T^{(\delta)}) \equiv 0, \qquad -\infty < \theta < \infty,$$

where U^* is the class of all nonrandomized unbiassed estimators of θ.

When the constant a is very large, the modification to the usual definition of variance apparently is very negligible, yet this slight change of variance completely wrecks the theory of unbiassed estimation. Not even locally best unbiassed estimators exist, let alone a uniformly best one.

In the construction of $T^{(\delta)}$, the independence of s^2 and $\bar{x}$ is not essential. As a matter of fact, we can replace s^2 by any real-valued statistic Y whose conditional

distribution, given $\bar{x}$, is continuous. We then replace c_δ by $c'_\delta(\bar{x})$, where $c'_\delta(\bar{x})$ is, say, the upper 100δ per cent point of the conditional distribution of Y given $\bar{x}$. From the sufficiency of $\bar{x}$ it follows that $c'_\delta(\bar{x})$ is independent of δ, and the rest of the proof follows through. Under similar circumstances the general theorem proved earlier will remain true in the restricted class U^* of all nonrandomized unbiassed estimators of μ_θ.

REFERENCES

[1] E. W. Barankin, "Locally best unbiased estimates," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 477–501.

[2] J. L. Hodges and E. L. Lehmann, "Some problems of minimax point estimation," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 182–197.

[3] E. L. Lehmann and H. Scheffé, "Completeness, similar regions, and unbiassed estimation," *Sankhyā*, Vol. 10 (1950), pp. 305–340.

[4] C. Stein, "Unbiased estimates with minimum variance," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 406–415.

ON CONFIDENCE INTERVALS OF GIVEN LENGTH FOR THE MEAN OF A NORMAL DISTRIBUTION WITH UNKNOWN VARIANCE

By Lionel Weiss[1]

University of Virginia

1. Summary. The problem of finding a confidence interval of preassigned length and of more than a given confidence coefficient for the unknown mean of a normal distribution with unknown variance is insoluble if the sample size used is fixed before sampling starts. In this paper two-sample plans, with the size of the second sample depending upon the observations in the first sample (as in [1]), are discussed. Consideration is limited to those schemes which increase the center of the final confidence interval by k if each observation is increased by k, and for which the size of the second sample is a function only of the differences among the observations in the first sample. Then it is shown that the mean of all the observations taken should be used as the center of the final confidence interval. Those schemes which make the size of the second sample a nondecreasing function of the sample variance of the first sample are shown to have certain desirable properties with respect to the distribution of the number of observations required to come to a decision.

2. Assumptions. We deal with an infinite sequence of independent and identically distributed chance variables, $(X_1, X_2, \cdots)$. Each has a normal distribution with unknown mean μ and unknown standard deviation σ. A positive integer $n \geqq 2$ is given, and $X_1, \cdots, X_n$ are observed. Then additional chance

---

Received October 21, 1953, revised September, 1954.

[1] Research under a grant from the Institute for Research in the Social Sciences, University of Virginia.

variables are observed to the number of $N(X_2 - X_1, X_3 - X_1, \cdots, X_n - X_1)$. This last expression is a measurable function of $X_2 - X_1, \cdots, X_n - X_1$, which we shall often abbreviate to N.

A fixed positive number Δ and a fixed β, with $0 < \beta < 1$, are given. We are going to study schemes that yield a confidence interval for μ of length Δ, and of confidence coefficient greater than β, regardless of the value of σ. We limit consideration to schemes such that after the second sample has been taken, the center of the final confidence interval is a function $C(X_1, \cdots, X_n, X_{n+1}, \cdots, X_{n+N})$ such that

$$C(X_1 + k, \cdots, X_{n+N} + k) = C(X_1, \cdots, X_{n+N}) + k,$$

identically in k and in those values of $X_1, \cdots, X_{n+N}$ for which we take a total of $n + N$ observations, for all values of N.

3. The optimum center of the final confidence interval. Once a function $N(X_2 - X_1, \cdots, X_n - X_1)$ has been assigned, then the distribution of the total sample size is determined. This distribution does not depend on μ for the class of schemes under consideration. Once we have chosen the function N, there is still the problem of assigning the center of the confidence interval, which will affect the confidence coefficient. (The confidence coefficient is a function of σ and of the function N, but not of μ for the cases we are discussing.) We have

LEMMA 3.1. *For any given function $N(X_2 - X_1, \cdots, X_n - X_1)$, setting the center of the confidence interval at the mean of all the $n + N$ observations taken maximizes the confidence coefficient uniformly in σ.*

PROOF. The joint conditional density of $X_1, \cdots, X_n, X_{n+1}, \cdots, X_{n+m}$, given that $N(X_2 - X_1, \cdots, X_n - X_1) = m$, is equal to

$$H(\sigma) \exp \{-\textstyle\sum_1^{n+m} (x_i - \mu)^2/2\sigma^2\}$$

in the region where $N = m$, and zero elsewhere. The exact form of $H(\sigma)$ is of no concern to us. Denoting by $\bar{X}$ the mean of the values $X_1, \cdots, X_{n+m}$, we can write $C(X_1, \cdots, X_{n+m})$ as $\bar{X} + G(X_1 - \bar{X}, \cdots, X_{n+m} - \bar{X})$, where the function G determines and is determined by the function C. A simple calculation shows that choosing $G(X_1 - \bar{X}, \cdots, X_{n+m} - \bar{X})$ identically equal to zero maximizes

$$P\{(\bar{X} + G - \tfrac{1}{2}\Delta) < \mu < (\bar{X} + G + \tfrac{1}{2}\Delta) \mid (N = m)\}$$

with respect to the function G for any value of σ, thus proving the lemma.

This calculation also shows that once we set $\bar{X}$ as the center of our confidence interval, the conditional probability, given that $N = m$, of having our confidence interval contain μ when σ is the standard deviation is equal to

$$\frac{\sqrt{(n + m)/2\pi}}{\sigma} \int_{-\Delta/2}^{\Delta/2} \exp \{-(n + m)t^2/2\sigma^2\}\, dt, \tag{1}$$

and does not depend on which function $N(X_2 - X_1, \cdots, X_n - X_1)$ is used.

We shall denote the quantity given in (1) by $\beta(\sigma, m)$. We note that $\beta(\sigma, m)$ is increasing in m for any fixed σ, and decreasing in σ for any fixed m. For fixed m, $\beta(\sigma, m)$ approaches unity as σ approaches zero.

4. Properties of a certain class of functions $N(X_2 - X_1, \cdots, X_n - X_1)$. We shall denote $\sum_1^n (X_i - \bar{X})^2$ by $S(X)$, with X standing for the generic point in $(X_1, \cdots, X_n)$ space. The symbol $P(R \mid \sigma)$ stands for the probability of the region R when the standard deviation has the value σ. We have

LEMMA 4.1. *Given any function $N(X_2 - X_1, \cdots, X_n - X_1)$, any non-negative integer m, and any positive value σ', there exists a non-negative number $d(\sigma', m)$ such that*

$$P[S(X) < d(\sigma', m) \mid \sigma] \gtreqless P[N(X_2 - X_1, \cdots, X_n - X_1) \leqq m \mid \sigma]$$

according to whether $\sigma \lesseqgtr \sigma'$.

PROOF. The two sets of values $(X_2 - X_1, \cdots, X_n - X_1)$ and $(X_2 - \bar{X}, \cdots, X_n - \bar{X})$ determine each other uniquely. The joint density of the second set is

$$p(X_2 - \bar{X}, \cdots, X_n - \bar{X}) = (K/\sigma^{n-1}) \exp\{-S(X)/2\sigma^2\}, \qquad K \text{ constant.}$$

From this, it follows easily that if $\sigma_0 < \sigma_1 < \sigma_2$, then of all regions R in $(X_2 - \bar{X}, \cdots, X_n - \bar{X})$ space with $P(R \mid \sigma_1)$ equal to a specified value, the region of the form $S(X) < d$, where d is a constant, minimizes $P(R \mid \sigma_2)$ and maximizes $P(R \mid \sigma_0)$. This proves the lemma.

5. Optimum properties of a certain class of procedures for setting the size of the second sample. For any sample-size rule R, $g(m, \sigma, R)$ denotes the probability that the size of the second sample will be m when the standard deviation is σ and R is used, while $G(m, \sigma, R)$ denotes the probability that the size of the second sample will not exceed m under the same circumstances. Let

$$B(\sigma, R) = \sum_{m=0}^{\infty} g(m, \sigma, R)\, \beta(\sigma, m), \qquad b(\sigma, m, R) = \sum_{i=0}^{m} g(i, \sigma, R)\, \beta(\sigma, i).$$

Thus $B(\sigma, R)$ is the confidence coefficient when R is used and σ is the standard deviation.

We shall say that a rule T setting the size of the second sample is of *type S* if it assigns size m to precisely those points $X = (X_1, \cdots, X_n)$ for which

$$a_{m-1}(T) < S(X) \leqq a_m(T), \qquad m = 0, 1, 2, \cdots,$$

where $a_{-1}(T)$, $a_0(T)$, $\cdots$ are preassigned numbers such that $0 = a_{-1}(T) \leqq a_0(T) \leqq a_1(T) \leqq \cdots$. It is clear that for any rule T of type S, if for some positive σ we have $G(m, \sigma, T)$ approaching unity as m increases, then this happens for all σ.

LEMMA 5.1. *If T is a sample-size rule of type S such that for any finite σ there is a finite positive integer $s(\sigma, T)$ such that $b[\sigma, s(\sigma, T), T] > \beta$, then for any σ' there is a finite positive integer $N(\sigma', T)$ such that $b[\sigma, N(\sigma', T), T] > \beta$ for all $\sigma \leqq \sigma'$.*

PROOF. From continuity considerations, it is easily seen that for any σ in the interval $(0, \sigma']$, if $b[\sigma, s(\sigma, T), T] > \beta$ then σ is in the interior of a nondegenerate

interval such that for any σ'' in the interval, $b[\sigma'', s(\sigma, T), T] > \beta$. By the Heine-Borel theorem, a finite number of such intervals can be found to cover the interval $[\delta, \sigma']$ for any positive δ. This proves the lemma, since a neighborhood of zero causes no trouble.

THEOREM 5.1. *Given any sample-size rule R such that $G(m, \sigma, R)$ approaches unity as m increases for any value of σ, and $B(\sigma, R) > \beta$ for all σ, there is a rule R' of type S, with $B(\sigma, R') \geqq \beta$ for all σ, such that for any σ there is a nonnegative finite integer $M(\sigma)$ so that*

$$G(m, \sigma, R') \geqq G(m, \sigma, R), \qquad \text{for all } m > M(\sigma).$$

PROOF. For the sake of definiteness, if T is any rule of type S, we shall understand that $N(\sigma', T)$ is the smallest positive integer such that $b[\sigma, N[\sigma', T], T] > \beta$ for all $\sigma \leqq \sigma'$.

Let σ_0 be the value of σ which is the solution to $\beta(\sigma, 0) = \beta$. Let δ denote a fixed positive number, and σ_i denote $\sigma_0 + i\delta$, for $i = 1, 2, \cdots$. Let R_0 be the uniquely determined rule of type S such that $g(m, \sigma_0, R_0) = g(m, \sigma_0, R)$ for all m. From the definition of σ_0, for any $\sigma \leqq \sigma_0$ there is a finite integer $s(\sigma, R_0)$ defined in Lemma 5.1.

Also, by Lemma 4.1, $G(m, \sigma, R_0) \leqq G(m, \sigma, R)$ for all $\sigma > \sigma_0$ and for all m. This fact and the assumptions about R imply that for each $\sigma > \sigma_0$ there is a finite $s(\sigma, R)$. We define N_1 as the larger of 1 or $N(\sigma_1, R_0)$. By Lemma 5.1, N_1 is finite.

Next we define another rule R_1 of type S, as follows. Take $a_i(R_1) = a_i(R_0)$ for $i = -1, \cdots, N_1 + 1$, while $a_i(R_1)$ for $i > N_1 + 1$ is chosen so that $G(i, \sigma_1, R_1) = G(i, \sigma_1, R)$. This is possible, since $G(N_1 + 2, \sigma_1, R) \geqq G(N_1 + 2, \sigma_1, R_0)$. It is clear that for any $\sigma \leqq \sigma_1$ there is a finite $s(\sigma, R_1)$. Also, by Lemma 4.1, $G(m, \sigma, R_1) \leqq G(m, \sigma, R)$ for any $\sigma > \sigma_1$. This implies that for any $\sigma > \sigma_1$, there is a finite $s(\sigma, R_1)$. Now we can define the finite number N_2 as the larger of $N_1 + 1$ or $N(\sigma_2, R_1)$.

Next we define a rule R_2 of type S, as follows. Take $a_i(R_2) = a_i(R_1)$ for $i = -1, 0, \cdots, N_2 + 1$, while $a_i(R_2)$ for $i > N_2 + 1$ is chosen so that $G(i, \sigma_2, R_2) = G(i, \sigma_2, R)$. By the same reasoning as above, we see that this is possible, and that we can define a finite $N(\sigma_3, R_2)$. Then we define N_3 as the larger of $N_2 + 1$ or $N(\sigma_3, R_2)$.

In general, once having defined the rule R_j of type S, we define a rule R_{j+1} of type S, as follows. Take $a_i(R_{j+1}) = a_i(R_j)$ for $i = -1, \cdots, N_{j+1} + 1$, while $a_i(R_{j+1})$ for $i > N_{j+1} + 1$ is chosen so that $G(i, \sigma_{j+1}, R_{j+1}) = G(i, \sigma_{j+1}, R)$. Then we define N_{j+2} as the larger of $N_{j+1} + 1$ or $N(\sigma_{j+2}, R_{j+1})$, and proceed as above.

Clearly, for any fixed i, $\lim_{j\to\infty} a_i(R_j)$ exists. As a matter of fact, above a certain j, $a_i(R_j)$ remains constant. Finally, we define a rule R' of type S by $a_i(R') = \lim_{j\to\infty} a_i(R_j)$. For each j, $B(\sigma, R_j) > \beta$ for any σ. Therefore, it is clear from continuity considerations that for any σ, $B(\sigma, R') \geqq \beta$. Also, for any given σ', there is a k so that $\sigma' < \sigma_k$.

Let m be any integer greater than $N_k + 1$. Then there is a nonnegative integer j so that $N_{k+j} + 1 < m \leqq N_{k+j+1} + 1$. Then for all σ, $G(m, \sigma, R_{k+j+1}) = G(m, \sigma, R')$. Also, $G(m, \sigma_{k+j}, R_{k+j}) = G(m, \sigma_{k+j}, R)$. Therefore by Lemma 4.1, $G(m, \sigma', R_{k+j}) \geqq G(m, \sigma', R)$. Now by construction, $G(m, \sigma, R_{k+j+1}) = G(m, \sigma, R_{k+j})$, for all σ. From these relations, we deduce that $G(m, \sigma', R') \geqq G(m, \sigma', R)$.This proves the theorem, for $M(\sigma')$ can be taken as $N_k + 1$.

REFERENCE

[1] Charles Stein, "A two-sample test for a linear hypothesis whose power is independent of the variance," *Ann. Math. Stat.*, Vol. 16 (1945), pp. 243–258.

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of interest

Personal Items

Max Astrachan was on leave of absence from his position as Professor of Statistics, USAF Institute of Technology, during July and August as Operations Analyst with Hq 49th Air Division in England.

Maurice H. Belz has been appointed to the recently created Chair of Statistics in the University of Melbourne.

Paul C. Clifford has been granted a leave of absence from Montclair State Teachers College to serve as Consultant on Quality Control for the Organization for European Economic Cooperation. This project, which is sponsored by the Foreign Operations Administration, will take Mr. Clifford to Norway, Sweden, Denmark, the Netherlands, Germany, France, Italy, and Turkey.

Mr. Gerald D. Berndt, formerly with the Quartermaster Research and Development Center at Natick, Mass., has accepted a special assignment with the Advisory Committee on Weather Control in Washington, D. C. The assignment is for 18 months and is concerned with analyzing meteorological experiments.

Dr. H. A. David has resigned from his position with the Commonwealth Scientific and Industrial Research Organization of Australia to take up a Senior Lectureship in Statistics in the University of Melbourne.

After receiving his Ph.D. from the Department of Mathematical Statistics of Columbia University, Cyrus Derman has joined the faculty of Syracuse University as an Instructor in the Department of Mathematics.

Frank M. duMas has joined the staff of the Department of Psychology, Montana State University. He formerly was a member of the faculty of Michigan State College.

John W. Forman, formerly an instructor in Mathematics at the University of Kansas, is now an Applied Science Representative for International Business Machines Corporation in Kansas City, Missouri.

Mr. Edwin L. Godfrey has accepted a position as Supervisory Mathematician in the Problem Formulation Section of the Statistical Services Division associated with the UNIVAC electronic computer at Wright-Patterson Air Force Base. He has just returned from New York where he was enrolled in a course in Programming conducted by the training division of Remington Rand, Inc.

Roe Goodman has been transferred from the Technical Assistance Administration of the United Nations to the Food and Agriculture Organization under whose program of technical assistance he has been sent to Chile in connection with the Chilean Agricultural Census, 1955. He and his family arrived in Santiago in September, 1954 for an assignment of one year.

Dr. T. N. E. Greville formerly a statistician with the United States Operations Mission to Brazil has been appointed Assistant Chief Actuary of the Social Security Administration in Washington, D. C.

Robert Hooke, of the Analytical Research Group, Princeton University, has accepted a position, starting in July, as a Research Mathematician with the Westinghouse Research Laboratories, East Pittsburgh, Pa.

Arthur Horberg, formerly of the Institute for Air Weapons Research, has accepted the position of Mathematician with the Bendix Research Laboratories, Detroit.

W. Robert Hydeman has been transferred by Remington Rand, Inc. to the Electronic Computer Department at 315 Fourth Ave., New York City. Mr. Hydeman's activities as Sales Engineer include the application of electronic computers to engineering and scientific calculations and the analysis of data handling systems employing electronic computers as their central processing facilities.

R. J. Jessen was on leave from the Statistical Laboratory, Iowa State College, for six weeks (from mid-February through March 1955) to serve as Technical Expert for FAO at a Latin American Demonstration Center on Agricultural Sampling Techniques being held in Buenos Aires, Argentina.

B. S. Kawar is now at the University of Illinois on a half-time teaching assistantship while working on his Ph.D. in Statistics.

Edward H. Kingsley is Operations Analyst in the Office of Operations Analysis, Deputy Chief of Staff/Operations, HQ, Air Proving Ground Command, Elgin Air Force Base, Florida.

Miss Evelyn L. Rumer has recently married and is now Mrs. Edward C. Kramer.

Upon completion of his Ph.D. in Statistics at Stanford University, Paul Meyer was drafted into the army. He is now stationed in Baltimore, at HQ-Materiel Command of the Chemical Corps, and is working as a statistician in the Quality Evaluation Division.

M. Ray Mickey has resigned from his position as Assistant Professor in the Statistical Laboratory of Iowa State College to rejoin Rand Corporation, Santa Monica, California, in April 1955. He will be part of a group working on logistic problems for the U. S. Department of Defense.

O. B. Moan has accepted a position as Research Specialist in the Staff Engineering Department of Lockheed Aircraft Corp., Missile Systems Division, Van Nuys, California.

Bruce D. Mudgett is doing a year of teaching at Brown University as Visiting Professor of Economics, filling in for a regular member of their staff who is spending a year with the National Bureau of Economic Research in New York.

Henry H. Peterson has accepted a position as Associate Engineer, Operations Analysis for the Douglas Aircraft Company at Long Beach, California.

Dr. Bayard Rankin has accepted a position as Research Associate in the Department of Mathematics, Massachusetts Institute of Technology.

Joseph S. Rhodes has resigned his position as Mathematical Statistician, Department of the Air Force, Washington, D. C. to accept a position as Methods and Operations Research Officer with the Chesapeake and Ohio Railway Company, Cleveland, Ohio.

Dr. W. A. Shewhart has been given the first Honorary Professorship in Statistical Quality Control by Rutgers University. He is at present in India and was this year elected to Honorary Membership in the Royal Statistical Society and awarded the Holley Medal of the American Society of Mechanical Engineers.

Rosedith Sitgreaves has accepted a position as Research Associate on Air Force Project on "Discriminatory Analysis" at Teachers College, Columbia University, New York.

Dr. Milton Sobel has accepted a position, starting Sept. 1, 1954, with the Bell Telephone Laboratories in Allentown, Pa. He was formerly working on an Air Force Research Contract at Cornell University.

Bergen R. Suydam has accepted the position of Chief, Operations Analysis, at Headquarters, United States Air Forces in Europe.

Kazuo Uemura has accepted the appointment by the World Health Organization of the United Nations to the post of Statistician in Geneva, Switzerland.

John Richard Wittenborn has been appointed at Rutgers University as University Professor of Psychology with assignment to the School of Education as Chairman of the Department of Pupil Personnel Service. This Department includes measurements and statistics, guidance, and school psychology.

In February Dr. E. S. Watson assumed the position of Senior Fellow in the Department of Statistics, Research School of the Social Sciences, Australian National University, Canberra, Australian Central Territory. The Australian National University, founded about four years ago is entirely a postgraduate university.

Martin B. Wilk, associate in the Statistical Laboratory and the Iowa Agricultural Experiment Station, has been awarded the George W. Snedecor Award in Statistics for 1955. The award, comprising a year's membership in the IMS and a subscription to its Annals, is given annually to the most outstanding candidate for the Ph.D. degree in statistics at Iowa State College.

Dr. Sewall Wright has retired from the University of Chicago (Department of Zoology) and has accepted the Leon J. Cole Professorship of Genetics at the University of Wisconsin.

Summer Offerings in Statistics at Iowa State College

The Department of Statistics at Iowa State College will offer a course in decision theory at the advanced graduate level during the first half of the 1955 summer quarter. The course will be taught by Dr. S. L. Isaacson. Members of the graduate faculty in statistics will be available during most of the summer for consultation on graduate research and for special problems courses. Other offerings for the two six-week sessions (June 13–July 20 and July 20–August 26) of the summer quarter are designed mainly for the graduate minor in statistics and for the beginning graduate major in statistics who wishes to satisfy prerequisite requirements for more advanced courses. These additional offerings include "Statistical Methods for Research Workers", "Statistical Theory for Research Workers", "Experimental Designs for Research Workers", and "Survey Designs

for Research Workers". For additional information, write to: T. A. Bancroft, Director, The Statistical Laboratory, Iowa State College, Ames, Iowa.

Summer Offerings in Statistics at the University of Michigan

During the Summer Session of 1955, June 20–August 12, the University of Michigan in Ann Arbor will offer a variety of programs and courses of interest to statisticians and students of mathematical statistics. The Department of Mathematics will present a number of elementary and advanced courses and seminars on theoretical statistics, probability, finite differences, and game theory. These courses will be under the direction of Professors A. H. Copeland, Sr., C. C. Craig, P. S. Dwyer, C. J. Nesbitt, and Dr. Clarke. A number of courses in applied statistics and statistical methods will also be given under the auspices of the Departments of Mathematics, Economics, Sociology, Psychology, and the School of Business Administration.

In addition to the preceding courses, a number of special programs in allied fields will be offered. Among these are the Program in Survey Research Techniques, July 18–August 12, the Summer Institute in Mathematics for Social Scientists, June 20–August 12, the Program on Digital Computers and Data Processors, August 1–12, and the Program on Quality Control, August 17–27.

Inquiries regarding courses in Statistics should be directed to Professor C. C. Craig, Department of Mathematics, and inquiries regarding the allied programs should be addressed to the Office of the Summer Session, 3510 Administration Building, Ann Arbor, Michigan.

Doctoral Dissertations in Statistics, 1954

Listed below are the doctorates conferred during the year 1954 in the United States and Canada for which the dissertations were written on topics in statistics or related fields. The university, major subject, and the title of the dissertation are given in each case.

S. G. Allen, Stanford, major in statistics, "A Class of Minimax Tests for One-sided Composite Hypotheses".

S. L. Andersen, North Carolina State College, major in experimental statistics, "Robust Tests for Variances and Effect of Non Normality and Variance Heterogeneity on Standard Tests".

G. E. Baxter, Minnesota, "An Application of Stochastic Processes to Certain Problems in the Brownian Motion of Continuous Media".

D. R. Bey, Illinois, major in statistics, "Modified Discriminant Analysis of School Organizations".

Allan Birnbaum, Columbia, major in mathematical statistics, "Characterizations of Complete Classes of Tests of Some Multiparametric Hypotheses with Applications to Likelihood Ratio Tests".

Isadore Blumen, North Carolina, major in statistics, "The Estimation of the Means of the Multivariate Normal Distribution: Minimax Solutions".

R. N. Bradt, Stanford, major in statistics, "On the Design and Comparison of Certain Dichotomous Experiments".

Barron Brainerd, Michigan, major in mathematics, "An Algebraic Theory of Probability with Applications to Analysis and Mathematical Logic".

Leo Breiman, California (Berkeley), major in mathematics, "Homogeneous Processes".

P. P. Crump, North Carolina State College, Major in experimental statistics, "Optimal Designs to Estimate the Parameters of a Variance Component Model".

M. B. Danford, North Carolina State College, major in experimental statistics, "Factor Analysis and Related Statistical Techniques".

Cyrus Derman, Columbia, major in mathematical statistics, "Some Contributions to the Theory of Markov Chains".

A. R. Eckler, Princeton, major in mathematics, "Rotation Sampling".

S. M. Free, Jr., North Carolina State College, major in experimental statistics, "Relationship of Instrument Color Measurements to Some Quality Indices of Flue-cured Tobacco".

W. R. Gaffey, California (Berkeley), major in mathematical statistics, "The Problem of Within-family Contagion".

R. K. Getoor, Michigan, major in mathematics, "Some Connections Between Operators in Hilbert Space and Random Functions of the Second Order".

A. H. E. Grandage, North Carolina State College, major in experimental statistics, "Biological Assay of a Material when Interfering Substances are Present".

L. E. Grosh, Purdue, major in mathematics, "Uniform Distribution and Round-off Error".

J. E. Hafstrom, Minnesota, major in mathematics, "Non-Linear Transformations in Wiener Space of the form, $y(t) = x(t) + q[t; x(t_1), \cdots, x(t_n)]$".

D. C. Hastings, Minnesota, major in statistics, "A Monthly Index of Regional Business Activity Based on Net Product Measurements".

J. S. Hunter, North Carolina State College, major in experimental statistics, "Multi-factor Experimental Designs".

P. K. Ito, St. Louis, "On the Simultaneous Minimax Point Estimation".

E. M. Jacobs, Iowa State College, major in statistics, "A Statistical Technique for Estimating the Characteristics of Consumer Behavior".

Maurice Kennedy, California Institute of Technology, "Ergodic Theorems for a Certain Class of Markoff Processes".

H. S. Konijn, California (Berkeley), major in mathematical statistics, "On the Power of Some Tests for Independence".

C. H. Kraft, California (Berkeley), major in mathematical statistics, "On the Problem of Consistent and Uniformly Consistent Statistical Procedures".

R. B. McHugh, Minnesota, major in statistics, "On the Scaling of Psychological Data by Latent Structure Analysis".

P. L. Meyer, Stanford, major in statistics, "An Application of the Invariance Principle to the Student Hypothesis".

T. V. Narayana, North Carolina, major in statistics, "Sequential Procedures in Probit Analysis".

K. C. S. Pillai, North Carolina, major in statistics, "On Some Distribution Problems in Multivariate Analysis".

J. H. Powell, Michigan State College, major in mathematical Statistics, "A Mathematical Model for Single Function Group Organization Theory with Applications to Sociometric Investigations".

Bayard Rankin, California (Berkeley), major in mathematical statistics, "The Concept of Sets Enchained by a Stochastic Process and its Use in Cascade Shower Theory".

W. C. Ross, Jr., State Univ. of Iowa, major in mathematics, "Certain Functions of Order Statistics".

Jacques Saint Pierre, North Carolina, major in statistics, "Distribution of Linear Contrasts of Order Statistics".

Herbert Scarf, Princeton, "Differential Operators on Manifolds, and Applications to Stochastic Processes".

R. F. Schweiker, Harvard, "Individual Space Models of Certain Statistics".

K. C. Seal, North Carolina, major in statistics, "On a Class of Decision Procedures for Ranking Means".

J. M. Shapiro, Minnesota, major in mathematics, "An Error Estimate for the Convergence of Distributions of Sums of Independent Random Variables to Infinitely Divisible Distributions".

Morris Skibinsky, North Carolina, major in statistics, "Some Properties of a Bayes Two-stage Test for the Mean".

Harry Smith, Jr., North Carolina State College, major in experimental statistics, "Weighting Coefficients for Age-adjusted Death Rates".

D. S. Stoller, California (Los Angeles), major in statistics, "Asymptotically Optimum Discriminant Function Estimates".

H. L. Stubbs, Boston, major in mathematics, "Non-normal Models for the Classification of Speech Sounds".

W. A. Thompson, North Carolina, major in statistics, "On the Ratio of Variances in the Mixed Incomplete Block Model".

A. W. Wortham, Oklahoma A and M, "On Estimates of Variance Components".

Z. S. Wurtele, Columbia, major in mathematical statistics, "Some Properties of Bayes Procedures which Improve Lot Quality".

New Members

The following persons have been elected to membership in the Institute

November 17, 1954 to February 8, 1955

Alder, Henry L., Ph.D. (Univ. of California), Assistant Professor of Mathematics, University of California, Davis, California.

Ansari, Abdur Rahman, M.A. (Muslim Univ., Aligarh, India), Graduate Assistant, Mathematics Department, University of Oregon, Eugene, Oregon.

Bell, Earl Louis, M.S. (Northwestern Univ.), Mathematical Statistician, Department of the Air Force, School of Aviation of Medicine, Department of Biometrics, Randolph Air Force Base, Texas, *483 Trudel Drive, San Antonio, Texas.*

Birdsall, T. G., M.S. (Univ. of Michigan), Research Associate, Engineering Research Institute, University of Michigan, Ann Arbor, Michigan.

Blasbalg, Herman, M.S.E.E. (Univ. of Maryland), Research Associate, Johns Hopkins University, 1315 St. Paul Street, Baltimore, Maryland.

Brown, William Fuller, Jr., Ph.D. (Columbia Univ.), Research Physicist, Sun Oil Company, Sun Physical Laboratory, Newtown Square, Pennsylvania, *600 Parrish Road, Swarthmore, Pennsylvania.*

Burros, Raymond H., Ph.D. (Yale Univ.), Visiting Associate Professor of Psychology, University of Houston, Houston, Texas.

Carroll, Mavis B. (Mrs.), B.Sc. (Rutgers Univ.), Project Leader—Statistician, Central Research Laboratories, General Foods Corporation, 11th and Hudson Sts., Hoboken, New Jersey.

Doyle, Patrick J., B.S. (Univ. of Washington), Mathematical Statistician, Naval Ordnance Laboratory, Corona, California.

Durbin, James, M.A. (Cambridge Univ.), Reader in Statistics, University of London, London School of Economics and Political Science, Houghton Street, London W.C.2, England.

Elias, Peter, Ph.D. (Harvard Univ.), Assistant Professor of Electrical Engineering, Massachusetts Institute of Technology, Cambridge 39, Massachusetts.

Garratt, Alfred E., M.S. (Northwestern Univ.), Chief, Design and Analysis Office, BW Assessment Laboratories, Dugway Proving Ground, Dugway, Utah, *145 North 2nd Street, Tooele, Utah.*

Golub, Gene H., M.A. (Univ. of Illinois), Research Assistant, Electronic Digital Computer, 162 Engineering Research Laboratory, University of Illinois, Urbana, Illinois.

Graves, Clayborn L., M.S. (Univ. of Southern Calif.), Supervisory Mathematical Statistician, Naval Ordnance Laboratory, Corona, California, *18382 E. Crescent Drive, Anaheim, California.*

Gregory, Geoffrey, B.A. (Cambridge Univ.), Research Assistant and Student, Department of Applied Mathematics and Statistics, Stanford University, Stanford, California.

Guha Thakurta, B. K., M.S. (Univ. of Calcutta), Assistant Superintendent of Statistics, State Statistical Bureau, Government of West Bengal, P.O.—Krishnagar, Dist.—Nadia, West Bengal, India.

Hall, W. B., B.A. (Univ. of Melbourne), Research Officer, Commonwealth Scientific and Industrial Research Organization, Division of Mathematical Statistics, Graham Road, Highett, S21, Victoria, Australia, *138 Whitehorse Road, Balwyn E.8., Victoria, Australia.*

Helmstadter, G. C., Ph.D. (Univ. of Minnesota), Associate in Research, Educational Testing Service, 20 Nassau Street, Princeton, New Jersey.

Homma, Tsuruchiyo, M.S. (Tôhoku Univ.), Lecturer of Mathematics, Mathematical Institute, Faculty of Science, Tokyo Metropolitan University, 591 Fusuma-cho, Meguro-ku, Tokyo, Japan.

Johnson, Millicent H., B.A. (Univ. of Minnesota), Teaching Assistant, Biostatistics Division, Department of Public Health, University of Minnesota, Minneapolis 14, Minnesota.

Katz, Morris W., B.A. (McMaster), Graduate Student in Statistics, University of Illinois, Urbana, Illinois, *92 Hillcrest Avenue, Hamilton, Ontario, Canada.*

Kennedy, George H., A.B. (Princeton Univ.), Analytical Statistician (Biological Sciences), U.S. Army Chemical Corps, Camp Detrick, Frederick, Maryland.

Lentz, Gordon A., B.A. (Univ. of Minnesota), Graduate Student, University of Minnesota, Minneapolis 14, Minnesota, *7508 3rd Avenue South, Minneapolis, Minnesota.*

Mardessich, Bartolo, Doctor in Statistics (Univ. of Rome), Assistant Director, DOXA—Institute for Statistical Research and Public Opinion Analysis, Via Mameli, 10—Milano, Italy, *Viale Lucania, 21—Milano, Italy.*

Miller, Irwin, M.S. (Purdue Univ.), Graduate Research Assistant. Department of Statistics, Virginia Polytechnic Institute, Blacksburg, Virginia.

Neff, M. Carolyn, B.S. (Western Reserve Univ.), Administrative Assistant, Development Division, Lecturer, Secretarial Sciences, School of Business, Graduate Student, School of Business, Western Reserve University, 2040 Adelbert Road, Cleveland 6, Ohio.

Oakford, Robert Vernon, B.A. (Stanford Univ.), Graduate Student in Statistics, Stanford University, Stanford, California, *10 Maywood Lane, Menlo Park, California.*

Perles, H. E., M.A. (Cantab), Director of Census and Statistics, Department of Census and Statistics, Government of Ceylon, P. O. Box 563, Colombo, Ceylon, *16, DeFonseka Road, Colombo 5, Ceylon.*

Pinkham, Roger S., A.M. (Harvard Univ.), Graduate Student, Harvard University, Cambridge, Massachusetts, *203C Holden Green, Cambridge, Massachusetts.*

Polowy, Henry, M.S. (New York Univ.), Instructor of Mathematics, Stevens Institute of Technology, Hoboken, New Jersey.

Robinson, William Keith, B.S. (Univ. of Washington), Actuary, Old American Life Company, 1619 Terry Avenue, Seattle 1, Washington.

Schlaifer, Robert, Ph.D. (Harvard Univ.), Associate Professor of Business Administration, Graduate School of Business Administration, Harvard University, Boston 63, Massachusetts.

Schneider, Robert C., B.S. (Seattle Univ.), Mathematical Statistician, Naval Ordnance Laboratory, Corona, California.

Shellow, James M., B.A. (Univ. of Chicago), Director of Research, Henry Shellow and Associates, 1012 North Third Street, Milwaukee, Wisconsin.

Shugert, Richard F., B.S. (Rutgers Univ.), Engineer Aerodynamicist, Applied Aeronautical Research, Stroukoff Aircraft Corporation, West Trenton, New Jersey.

Silvey, S. D., Ph.D. (Glasgow Univ.), Lecturer in Mathematics, University of Glasgow, Glasgow W.2., Scotland.

Star, Martin L., B.B.A. (City College of New York), Quality Control Supervisor, Sonotone Corporation, Cold Spring, New York, *2270 81st Street, Brooklyn 14, New York.*

Stiver, Ray B., Jr., M.B.A. (Univ. of Buffalo), Dynamics Engineer, Bell Aircraft Corporation, Wheatfield Plant, P.O. Box 1, Buffalo 5, New York, *914 Michigan Avenue, Buffalo 3, New York.*

Troll, John H., M.A. (Harvard Univ.), Vice President for Engineering, Electronics Corporation of America, 77 Broadway, Cambridge 42, Massachusetts.

Udinski, W. P., Ph.D. (Univ. of Illinois), Mathematician, Human Resources Research and Personal Training Section, Machine Processing Branch, Lackland Air Force Base, San Antonio, Texas, *714 South Palmetto Avenue, San Antonio 3, Texas.*

Yarnold, James K., B.S. (Trinity Univ.), Research Assistant, Statistical Laboratory, University of California, Berkeley 4, California.

Yasuoka, Yoshinori, M.Sc. (Tokyo Univ.), Lecturer of Mathematics, Mathematical Institute, Faculty of Science, Tokyo Metropolitan University, 591 Fusuma-cho, Meguro-ku, Tokyo, Japan, *2-242 Okuzawa-cho, Setagaya-ku, Tokyo, Japan.*

REPORT OF THE TREASURER FOR 1954

The report of the treasurer has been delayed this year on the recommendation of the treasurer and the concurrence of the Executive Committee in order that the report of the auditor might be submitted to the membership along with that of the treasurer. Because of the timing of publication of the *Annals*, it has been possible to give financial statements which more accurately reflect the position of the Institute on December 31, 1954 without introducing any additional delay in the publication of the reports. In previous years the books of record of the Institute have been closed in time to prepare financial reports for presentation at the Annual Meeting. While Statements of Income and Expense accurately reflected the activities through the whole of the calendar year, the Statements of Financial Condition did not include the cash receipts and the associated liabilities to members and subscribers which accumulated during the latter part of December. The Statement of Income and Expense submitted herewith can properly be compared with those for the last few years. The Statement of Financial Condition will differ in the three accounts mentioned in part because of the change in time of preparation.

Income and expenses are both higher in 1954 than in 1953. The increase in surplus is much lower in 1954 than in the two preceding years. Expenses increased for several reasons. One major increase was due primarily to the increase from 708 numbered pages in the *Annals* in 1953 to 826 pages in 1954. Along with an increase in the number of copies distributed, this item accounts for a $2,141.94 increase in expenses. Printing costs are rising and the cost per page will increase markedly in 1955. Another major increase in expenses resulted from the decision of the Council that at least twice a year the Executive Committee should meet in person and that, to assure these meetings, the Institute should pay travel expenses of members of the Executive Committee to the Summer Meeting and to the Annual Meeting each year. This item amounted to $990.98 in 1954. Other factors in the increase in expenses were the larger volume of material mailed to members by the Secretary and the use of Air Mail in distributing ballots on two occasions during the year to members outside the United States and Canada.

The decrease in dues rate for members outside the United States and Canada which became effective in 1954 would have produced a decrease of $410 in income if no new members had been accepted. Increases in membership and reinstatements reduced the decrease to $186. An increase of $354 from individual members in the United States and Canada (in spite of a reduction in student dues rate), $300 from institutional members and $828.35 from nonmember subscriptions to the *Annals* gave a net increase of $1,296.35 from dues and subscriptions. Changes in the form of investments and increased investments yielded $1,350.18, an increase of $1,025.18 over 1953.

The financial reports follow.

THE INSTITUTE OF MATHEMATICAL STATISTICS

Statement of Financial Condition
December 31, 1954

ASSETS

Current Assets

Cash in Bank		$10,689.37
Savings Bank Account		2,058.98
Investments (at cost)*		43,771.51
Dues Receivable		73.00
Due on Back Issue Sales		1,107.20
Miscellaneous Accounts Receivable		104.65
Inventory of Back Issues		18,524.28
Total Assets		$76,328.99

LIABILITIES AND MEMBERS EQUITY

Current Liabilities

Account Payable, Printing of Dec. *Annals* (Estimate)	$3,220.32	
Withholding and F.I.C.A. taxes payable	187.20	
Accrued Expenses	531.05	
Amount Held for *Biometrika* subscriptions	404.50	
Amount Held for Wald Memorial volume	1,440.00	
Total Current Liabilities		5,783.07

Liabilities to Members and Subscribers

Advances on Dues, 1955, 1956	$5,851.31	
Advance on Dues, life members	2,872.50	
Advances on Subscriptions	5,707.05	
Total Liabilities to Members and Subscribers		14,430.86

Members Equity

Reserve for Maintaining Supply of Back Issues	$21,752.52	
Surplus	34,362.54	
Total Members Equity		56,115.06
Total Liabilities and Members Equity		76,328.99

Statement of Income and Expenses
for the period January 1, 1954 through December 31, 1954

Revenues

Membership Dues	$14,115.00	
Subscriptions	10,071.00	

* The book value of investments is cost. The cash value is $6 higher. During 1954, $20,000 was transferred from Time Certificates of Deposit to Savings and Loan Associations. Investments at the end of 1954 were $11,888 in U. S. government bonds and $31,883.52 in Savings and Loan Associations. These amounts are fully insured by Agencies of the U. S. government.

Sale of Back Issues:			
Gross Income	$5,498.83		
Cost of Sales	1,884.04	3,614.79	
Income from Investments		1,350.18	
Miscellaneous Revenue		534.33	
Total Revenue			29,685.30
Expenses			
Printing of Current *Annals*:			
Cost of current issues printed	$13,927.98		
Inventory, December 31, 1954 of 1954 *Annals*	1,448.13	12,479.85	
Editorial Expense		849.69	
Miscellaneous Printing, Stationery, Postage		2,873.40	
Miscellaneous Office Expense		755.60	
Salary		3,360.00	
Contributions		151.73	
President's Office Expenses		131.00	
Travel		990.98	
F. I. C. A. tax		33.60	
Total Expenses			$21,625.85
Excess of Revenues over Expenses			8,059.45
Minus Addition to Reserve for Maintaining Supply of Back Issues			1,323.00
Increase in Surplus			$6,736.45
Surplus, December 31, 1953			$27,626.09
Surplus, December 31, 1954			$34,362.54

K. J. ARNOLD
Treasurer

I have examined the financial statements of the Institute of Mathematical Statistics as of December 31, 1954 consisting of a Statement of Financial Condition as of this date and a Statement of Revenue and Expense for the year then ended.

My examination was made in accordance with generally accepted auditing standards and included such tests of the accounting records as were considered necessary. The inventory of back issues was not physically examined, but direct confirmation was received from the independent agent who is responsible for the physical control of such inventory. Accounts receivable were not confirmed; however, they were not considered to be of a material amount.

In my opinion, the accompanying Statement of Financial Position and the Revenue and Expense statement present fairly the financial position of the Institute of Mathematical Statistics as of December 31, 1954 and the results of

its operations for the year then ended, in conformity with generally accepted accounting principles applied on a basis consistent with that of the preceding year.

CHARLES LAWRENCE
CPA (Michigan)

PUBLICATIONS RECEIVED

JONES, B. W., *The Theory of Numbers*, Rinehart and Co., Inc., New York, 1955, xi, 143 pp., $3.75.

Selected Papers in Statistics and Probability by Abraham Wald, McGraw-Hill Book Co., Inc., New York, 1955, ix + 702 pp., $8.00.

THEIL, H., *Linear Aggregation of Economic Relations*, North-Holland Publishing Co., Amsterdam, 1954, xi + 205 pp., $5.00.

Tinbergen, J., *Centralization and Decentralization in Economic Policy*, North-Holland Publishing Co., Amsterdam, 1954, viii + 80 pp., $1.80.

THE INSTITUTE [illegible] STATISTICS

President:
Henry [illegible], California
President-Elect:
[illegible], California
Secretary:
[illegible], East Lansing
[illegible], North Carolina
Treasurer:
To [illegible] State College, East Lansing
From July 1, [illegible] University, Stanford, Calif.
Editor:
E. L. Lehmann, [illegible]

The purpose of the [illegible] Statistics [illegible] the development, dissemination [illegible] mathematical [illegible]

Membership dues [illegible] the ANNALS OF MATHEMATICAL STATISTICS [illegible] residents of the United States or Canada [illegible] countries. There [illegible] rates for [illegible] for [illegible] members. Inquiries regarding membership in the Institute [illegible] the Secretary of the Institute.

MEETINGS OF THE INSTITUTE

WESTERN REGIONAL [illegible], CALIFORNIA

To be held in [illegible]

[illegible] MEETING [illegible]

To be held [illegible] American Mathematical Society [illegible] Mathematical [illegible] Society

[illegible] ANNUAL MEETING [illegible], NEW YORK

To be held in connection [illegible] Association and other [illegible]

(Abstracts [illegible] at least [illegible] days before the [illegible] presented. Abstract blanks may be [illegible])